

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2004, 08:55:46 ; Search time 53 Seconds  
(without alignments)  
421.156 Million cell updates/sec

Title: US-09-972-032-2  
Perfect score: 456  
Sequence: 1 MCGPRRVSACGFADAHWT.....SAGLTVRDPLQLGCMGRG 79

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq 29Jan04: \*  
1: geneseqp1980s: \*  
2: geneseqp1990s: \*  
3: geneseqp2000s: \*  
4: geneseqp2001s: \*  
5: geneseqp2002s: \*  
6: geneseqp2003as: \*  
7: geneseqp2003bs: \*  
8: geneseqp2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	456	100.0	79	AB080035	AB080035 Human ERC
2	82.5	18.1	168	ABR56758	ABR56758 Human sec
3	79.5	17.4	151	ABG22219	ABG22219 Novel hum
4	78	17.1	11096	AA310129	AA310129 Streptomy
5	77.5	17.0	426	ABG16164	ABG16164 Novel hum
6	77	16.9	215	ADA54191	ADA54191 Human pro
7	77	16.9	235	ABP95644	ABP95644 Human GPC
8	76.5	16.8	206	ABG14023	ABG14023 Novel hum
9	75.5	16.6	443	ABG03615	ABG03615 Novel hum
10	75.5	16.6	1063	ADC86843	ADC86843 Human GPC
11	75.5	16.6	1356	ADC86885	ADC86885 Human GPC
12	74.5	16.3	104	ABM65289	ABM65289 Propionib
13	74.5	16.3	156	ABP69309	ABP69309 Human pol
14	74.5	16.3	180	ABG14843	ABG14843 Novel hum
15	74.5	16.3	180	ABG12827	ABG12827 Novel hum
16	74.5	16.3	329	AAO10468	AAO10468 Human pol
17	74	16.2	393	ABG97894	ABG97894 Human sec
18	73.5	16.1	1180	ADC30896	ADC30896 Human nov
19	72.5	15.9	172	AAU65951	AAU65951 Propionib
20	72.5	15.9	172	ABM62270	ABM62270 Propionib
21	72.5	15.9	372	ADC07760	ADC07760 Rice prot
22	72	15.8	371	ABG03350	ABG03350 Novel hum
23	72	15.8	384	ABG06095	ABG06095 Novel hum
24	72	15.8	514	AAW25917	AAW25917 Human pro
25	71.5	15.7	421	ABG19501	ABG19501 Novel hum

26	71	15.6	93	4	AAU45659	AAU45659 Propionib
27	71	15.6	93	6	ABM42178	ABM42178 Propionib
28	71	15.6	192	4	ABG26053	ABG26053 Novel hum
29	71	15.6	253	4	AAE03237	AAE03237 Human gen
30	71	15.6	253	5	ABG64383	ABG64383 Human alb
31	71	15.6	359	4	ABB60902	ABB60902 Drosophil
32	70.5	15.5	638	2	AAU04992	AAU04992 Mycobacte
33	70	15.4	532	2	AAW40114	AAW40114 Human alp
34	70	15.4	1214	4	ABG61851	ABG61851 Drosophil
35	70	15.4	1669	4	AAW40863	AAW40863 Human pol
36	70	15.4	1669	5	ABB90760	ABB90760 Human tum
37	70	15.4	1672	4	ABU54467	ABU54467 Human tum
38	70	15.4	1672	4	AAW39077	AAW39077 Human pol
39	69.5	15.2	170	2	AAU29195	AAU29195 Amino aci
40	69.5	15.2	552	4	AAW55591	AAW55591 Human aci
41	69	15.1	101	3	AAW93886	AAW93886 Amino aci
42	69	15.1	19938	6	ABP76681	ABP76681 Streptomy
43	68.5	15.0	207	5	ABJ10911	ABJ10911 Human sec
44	68.5	15.0	207	7	ADB64597	ADB64597 Human pro
45	68.5	15.0	591	4	ABG03618	ABG03618 Novel hum

## ALIGNMENTS

RESULT 1  
AB080035  
ID ABB080035 standard; protein; 79 AA.  
XX  
AC ABB080035;  
XX  
DT 27-AUG-2002 (first entry)  
XX  
DE Human ERCoA3 protein.  
XX  
KW Estrogen Receptor Coregulator 3; ERCoA3; tamoxifen; estrogen; cancer;  
KW osteoporosis; cytostatic; osteopathic; human; receptor.  
XX  
OS Homo sapiens.  
XX  
FN W0200228352-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 05-OCT-2001; 2001WO-US031271.  
XX  
PR 05-OCT-2000; 2000US-0238190P.  
XX (UYCA-) UNIV CASE WESTERN RESERVE.  
XX Montano M, Sutton A;  
XX WPI; 2002-454492/48.  
XX N-PSDB; ABL06066.  
XX  
XX New polypeptide, that is a functional equivalent of ERCoA3 (Estrogen Receptor Coregulator 3), is useful in inhibiting or reducing tamoxifen or estrogen-induced proliferation of cancer cells and in treating osteoporosis.  
XX  
XX Claim 1; Fig 2; 39pp; English.  
XX  
XX The invention relates to a ERCoA3 (Estrogen Receptor Coregulator 3) protein and encoding polynucleotides. ERCoA3 can be used to inhibit or reduce tamoxifen or estrogen induced proliferation of cancer cells, by reduced activity of ERCoA3, and for detecting cancer cells that are tamoxifen resistant, or to treat osteoporosis, by increasing levels of ERCoA3 in cells. The encoding polynucleotide can be used to inhibit ERCoA3 translation of a mRNA encoding ERCoA3. ERCoA3 acts as a coregulator protein and can bind to the estrogen receptor to activate a molecular or cellular response in the cell. The present sequence represents the human ERCoA3 protein

SQ Sequence 79 AA;  
 Query Match 100.0%; Score 456; DB 5; Length 79;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-42;  
 Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCGRRPRVSAGCGFADAHWTGLTGLGEGGIGPEGQASPTDCA SRWPSASRPWS 60  
 DB 1 MCGRRPRVSAGCGFADAHWTGLTGLGEGGIGPEGQASPTDCA SRWPSASRPWS 60  
 QY 61 AGLTVDRPQLGELCMRG 79  
 DB 61 AGLTVDRPQLGELCMRG 79

RESULT 2  
 ABR56758  
 ID ABR56758 standard; protein; 168 AA.  
 XX ABR56758;  
 XX  
 DT 30-JUL-2003 (first entry)  
 XX  
 DE Human secreted protein SECP-33 SEQ ID NO:33.  
 XX  
 KW Human; secreted protein; SECP; cytostatic; antiarteriosclerotic;  
 KW anticonvulsant; neurotropic; neuroprotective; cerebroprotective; anti-HIV;  
 KW anti-allergic; anti-inflammatory; thymomimetic; gene therapy; cancer;  
 KW cell proliferative disorder; atherosclerosis; neurological disorder;  
 KW epilepsy; Huntington's disease; stroke; immune disorder; allergy; AIDS;  
 KW inflammatory disorder; developmental disorder; hypothyroidism;  
 KW Cushing's syndrome; infection.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO2003016506-A2.  
 XX  
 PD 27-FEB-2003.  
 XX  
 PF 15-AUG-2002; 2002WO-US027143.  
 XX  
 PR 17-AUG-2001; 2001US-0313249P.  
 PR 24-AUG-2001; 2001US-0314752P.  
 PR 07-SEP-2001; 2001US-0317818P.  
 PR 07-SEP-2001; 2001US-0317824P.  
 PR 21-SEP-2001; 2001US-0324040P.  
 PR 24-SEP-2001; 2001US-0324586P.  
 PR 02-NOV-2001; 2001US-0343980P.  
 PR 28-NOV-2001; 2001US-0334229P.  
 PR 13-FEB-2002; 2002US-0357002P.  
 PR 06-MAR-2002; 2002US-0362439P.  
 PR 19-MAR-2002; 2002US-0366041P.  
 PR 30-APR-2002; 2002US-0376988P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Tang TY, Warren BA, Gietzen KJ, Lal PG, Yue H, Honchell CD;  
 PI Lehr-Mason PM, Burford N, Xu Y, Baughn MR, Duggan BM, Tran UK;  
 PI Lee EA, Forsythe LJ, Richardson TW, Lee S, Thangavelu K, Yue H;  
 PI Emerling BM, Wallia NK, Azimzai Y, Sanjanwala B, Hafalia AJA;  
 PI Borowsky ML, Nguyen DB, Ison CH, Astromoff A, Ding L, Lee SY;  
 PI Becha SD, Ramkumar J, Gandhi AS, Jin P, Fu GK, Swarnakar A;  
 XX  
 DR WPI; 2003-278569/27.  
 DR N-PSDB; ACC79058.  
 XX  
 PT New human secreted proteins (SECP), useful for diagnosing, treating and  
 PT preventing diseases or conditions associated with the aberrant SECP  
 PT expression e.g. cancer, AIDS, atherosclerosis, Huntington's disease,  
 PT stroke, infections.  
 XX  
 PS Claim 1; Page 215; 286pp; English.  
 XX

CC ACC79026 to ACC79105 encode the human secreted proteins (I) given in  
 CC ABR56726 to ABR56805, designated SECP-1 to SECP-80. SECP sequences can  
 CC have cytostatic, antiarteriosclerotic, anticonvulsant, anti-inflammatory,  
 CC neurotropic, neuroprotective, cerebroprotective, anti-HIV, anti-allergic and  
 CC thymomimetic activities, and can be used in gene therapy. The SECP  
 CC proteins and polynucleotides can be used in diagnosing, treating and  
 CC preventing diseases or conditions associated with the decreased  
 CC expression or overexpression of SECP, such as cell proliferative (e.g.  
 CC cancer, atherosclerosis), neurological (e.g. epilepsy, Huntington's  
 CC disease, stroke), immune/inflammatory (e.g. AIDS, allergies) and  
 CC developmental (e.g. hypothyroidism, Cushing's syndrome) disorders, or  
 CC infections. They are also useful in assessing the effects of exogenous  
 CC compounds on the expression of nucleic acid and amino acid sequences of  
 CC SECP. The SECP or its fragments are useful in screening compounds for  
 CC effectiveness as agonist or antagonist of the polypeptides, or in  
 CC altering the expression of the target polynucleotide and compounds that  
 CC specifically bind to or modulate the activity of the polypeptide  
 XX  
 XX Sequence 168 AA;  
 SQ

Query Match 18.1%; Score 82.5; DB 6; Length 168;  
 Best Local Similarity 32.1%; Pred. No. 0.43;  
 Matches 25; Conservative 4; Mismatches 36; Indels 13; Gaps 4;

QY 3 GRPR---RVSAGCGFADAHWTGLTGLGEGGIGPEGQASPTDCA SRWPSASRPW 59  
 DB 81 GQPRGNPAQEGCGHVDGS-SLRWGLGPGSHGG-----KKWPPPLPPRPRG---WPP 130  
 QY 60 SAGLTVDRPQLGELCMG 77  
 DB 131 SQVAQVRLPREDRCSG 148

RESULT 3  
 ABR22219  
 ID ABR22219 standard; protein; 151 AA.  
 XX  
 AC ABR22219;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #22210.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US008631.  
 XX  
 PR 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS86406.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX  
 PS Claim 20; SEQ ID NO 52578; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain

CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG0377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 151 AA;

Query Match 17.4%; Score 79.5; DB 4; Length 151;  
 Best Local Similarity 37.0%; Pred. No. 0.83;  
 Matches 17; Conservative 6; Mismatches 14; Indels 9; Gaps 2;

Qy 27 GGGGEGGIGPGQA-----SPTDCASRWPRSRWFSAGLTVR 66

Db 55 GAGREHGMGPGCRAQPLKUSKSPGPKC---WVEQMRHWTTQLSLK 97

RESULT 4

AAE10129  
 ID AAE10129 standard; protein; 11096 AA.

AC AAE10129;

XX 29-NOV-2001 (first entry)

XX Streptomyces noursei nystatin gene, Nysc.

XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;  
 KW antifungal; antibiotic; PKS type I.

OS Streptomyces noursei.

XX Key Location/Qualifiers

FT Domain 35..455

FT /label= KS3 domain

FT /note= "Ketosynthase (KS) domain"

FT Domain 546..858

FT /label= AT3 domain

FT /note= "Acyltransferase (AT) domain"

FT Domain 872..1073

FT /label= DH3 domain

FT /note= "Dehydratase (DH) domain"

FT Domain 1381..1628

FT /label= KR3 domain

FT /note= "Ketoreductase (KR) domain"

FT Domain 1662..1735

FT /label= ACP3 domain

FT /note= "Acyl carrier protein (ACP) domain"

FT Domain 1757..2180

FT /label= KS4 domain

FT /note= "Ketosynthase (KS) domain"

FT Domain 2291..2603

FT /label= AT4 domain

FT /note= "Acyltransferase (AT) domain"

FT Domain 2617..2818

FT /label= DH4 domain

FT /note= "Dehydratase (DH) domain"

FT Domain 3124..3371

FT /label= KR4 domain

FT

FT Domain /note= "Ketoreductase (KR) domain"  
 FT 3407..3480  
 FT /label= ACP4 domain  
 FT /note= "Acyl carrier protein (ACP) domain"  
 FT Domain 3501..3924  
 FT /label= KS5 domain  
 FT /note= "Ketosynthase (KS) domain"  
 FT 4032..4346  
 FT /label= AT5 domain  
 FT /note= "Acyltransferase (AT) domain"  
 FT Domain 4360..4561  
 FT /label= DH5 domain  
 FT /note= "Dehydratase (DH) domain"  
 FT 4953..5239  
 FT /label= ER5 domain  
 FT /note= "Enoylreductase (ER) domain"  
 FT 5248..5495  
 FT /label= KR5 domain  
 FT /note= "Ketoreductase (KR) domain"  
 FT 5528..5601  
 FT /label= ACP5 domain  
 FT /note= "Acyl carrier protein (ACP) domain"  
 FT 5623..6046  
 FT /label= KS6 domain  
 FT /note= "Ketosynthase (KS) domain"  
 FT 6166..6478  
 FT /label= AT6 domain  
 FT /note= "Acyltransferase (AT) domain"  
 FT 6492..6704  
 FT /label= DH6 domain  
 FT /note= "Dehydratase (DH) domain"  
 FT 7038..7281  
 FT /label= KR6 domain  
 FT /note= "Ketoreductase (KR) domain"  
 FT 7315..7388  
 FT /label= ACP6 domain  
 FT /note= "Acyl carrier protein (ACP) domain"  
 FT 7408..7831  
 FT /label= KS7 domain  
 FT /note= "Ketosynthase (KS) domain"  
 FT 7939..8253  
 FT /label= AT7 domain  
 FT /note= "Acyltransferase (AT) domain"  
 FT 8267..8470  
 FT /label= DH7 domain  
 FT /note= "Dehydratase (DH) domain"  
 FT 8812..9086  
 FT /label= KR7 domain  
 FT /note= "Ketoreductase (KR) domain"  
 FT 9120..9193  
 FT /label= ACP7 domain  
 FT /note= "Acyl carrier protein (ACP) domain"  
 FT 9214..9637  
 FT /label= KS8 domain  
 FT /note= "Ketosynthase (KS) domain"  
 FT 9758..10072  
 FT /label= AT8 domain  
 FT /note= "Acyltransferase (AT) domain"  
 FT 10086..10289  
 FT /label= DH8 domain  
 FT /note= "Dehydratase (DH) domain"  
 FT 10657..10904  
 FT /label= KR8 domain  
 FT /note= "Ketoreductase (KR) domain"  
 FT 10939..11012  
 FT /label= ACP8 domain  
 FT /note= "Acyl carrier protein (ACP) domain"

WO200159126-A2.

16-AUG-2001.

08-FEB-2001; 2001WO-GB000509.

```

XX 08-FEB-2000; 2000GB-00002840.
PR 10-APR-2000; 2000GB-00008786.
PR 14-APR-2000; 2000GB-00009387.
XX
PA (UNNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA (SNTF) SINTEF STIFTELSEN IND TEK FORSK.
PA (ALPH-) ALPHARMA AS.
PA (SINV-) SINVENT AS.
PA (DZIE/) DZIEGLEWSKA H.
PA (ZORC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (FJAE/) FJAEVRIK E.
PA (BRAU/) BRAUTASET T.
PA (STRO/) STROM A R.
PA (VALL/) VALLA S.
XX
PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
DR WPI; 2001-557614/62.
DR N-PSDB; AAD17184.
XX
PT New nystatin polyketide synthase polynucleotides and polypeptides, useful
PT as antibiotics and antifungals.
XX
PS Claim 15; Page 170-176; 266pp; English.
XX
CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrolide antibiotic nystatin. The
CC nystatin PKS is useful as antifungal antibiotics. The present sequence is
CC a PKS type I encoding Streptomyces noursei nystatin gene, Nysc
XX
SQ Sequence 11096 AA;
Query Match 17.1%; Score 78; DB 4; Length 11096;
Best Local Similarity 37.2%; Pred. No. 1.3e+02;
Matches 32; Conservative 7; Mismatches 33; Indels 14; Gaps 6;
Qy 4 RPRVSVAGCGPADAHWTGL---WTGLGEGGEGGIGPEGQASPT-PDCASR-WPRS--ASR 56
Db 836 RPELSAVTGLARHVRGVTVRWAGLFD---GTGARRADLTPTYPFHQRFWPTAARAAQ 891
Qy 57 WPSAGLTVDRDRPOLG---ELCMGRG 79
Db 892 DVTAAAGLGAADHPULGATVELADGAG 917
XX
RESULT 5
ID ABG16164 standard; protein; 426 AA.
XX
AC ABG16164;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #16155.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX

```

---

```

PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS80351.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 46523; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostic, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ASG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: the sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 426 AA;
Query Match 17.0%; Score 77.5; DB 4; Length 426;
Best Local Similarity 37.0%; Pred. No. 4.2;
Matches 17; Conservative 2; Mismatches 18; Indels 9; Gaps 1;
Qy 36 PRQVSPPTDCASRWP-----RSASRWPWSAGLTVDRDRPOLG 72
Db 134 PPHLPSPDGLQGPSRIVGVSSEGEWFWQASLQVRCGRHILG 239
XX
RESULT 6
ADA54191
ID ADA54191 standard; protein; 215 AA.
XX
AC ADA54191;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human protein, SEQ ID 1759.
XX
KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease.
XX
OS Homo sapiens.
XX
PN EP1293569-A2.
XX
PD 19-MAR-2003.
XX
PF 21-MAR-2002; 2002EP-00006586.
XX
PR 14-SEP-2001; 2001JP-00328381.
PR 24-JAN-2002; 2002US-0350435P.
XX
PA (HELI-) HELIX RES INST.

```

(REAS-) RES ASSOC BIOTECHNOLOGY.

PA Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;  
 XX  
 DR WPI; 2003-395539/38.  
 DR N-PSDB; ADA52552.  
 XX  
 PT New polynucleotides encoding full-length polypeptides, e.g. secretory  
 PT and/or membrane proteins, useful for developing medicines for diseases in  
 PT which the gene is involved, or as target molecules for gene therapy.  
 XX  
 PS Claim 14; SEQ ID NO 1759; 205pp; English.  
 XX  
 CC The present invention relates to novel human secretory or membrane  
 CC proteins (ADA54072-ADA54710) and their coding sequences (ADA52433-  
 CC ADA54071). The coding sequences are useful in the gene therapy of  
 CC diseases caused by abnormalities of the proteins, e.g. cancer,  
 CC inflammatory diseases, osteoporosis or neurological disease.  
 XX  
 SQ Sequence 215 AA;

Query Match 16.9%; Score 77; DB 6; Length 215;  
 Best Local Similarity 31.9%; Pred. No. 2.3;  
 Matches 23; Conservative 3; Mismatches 28; Indels 18; Gaps 4;

QY 7 RVSGCGFADAHWTGL-----WTGLGEGGEGGIGPEGOASPTPDC-ASRWPRSRASRW 57  
 DB 64 RCGARC-----WEGSSLAQLQPPWPSGSLTAAGVQVCAD---SCWRSWPSRSPRW 114  
 QY 58 FWSAGLTVDRRP 69  
 DB 115 PPSSTFAARSSP 126

RESULT 7  
 ID ABP95644 standard; protein; 235 AA.  
 AC ABP95644;  
 XX  
 DT 06-MAR-2003 (first entry)  
 DE Human GPCR polypeptide SEQ ID NO 98.  
 KW Human; GPCR; G protein coupled receptor; signal transduction; olfactory;  
 KW drug development; gustatory; taste; fragrance; receptor.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200216548-A2.  
 XX  
 PD 28-FEB-2002.  
 XX  
 PF 30-JUL-2001; 2001WO-IB001446.  
 XX  
 PR 04-AUG-2000; 2000JP-00237818.  
 PR 13-FEB-2001; 2001JP-00034434.  
 XX  
 PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
 XX  
 PI Haga T, Takeda S, Mitaku S;  
 XX  
 DR WPI; 2002-304119/34.  
 DR N-PSDB; AB242918.  
 XX  
 PT Database global search for G protein-coupled receptors, proteins and  
 PT encoded genes for studying in vivo signal transduction mechanism and  
 PT identifying targets for drug development.  
 XX  
 PS Claim 10; SEQ ID NO 98; 97pp + Sequence Listing; Japanese.  
 XX

CC The invention relates to a method for screening G protein-coupled  
 CC receptor (GPCR) genes (ABZ42870-ABZ43216) and/or GPCR proteins (ABP95596-  
 CC ABP95942) by extracting open-reading frames containing 6-8 transmembrane  
 CC domains with 250-1000 amino acid residues to give a gene homologous with  
 CC a known GPCR gene. The receptor proteins and encoded genes are useful for  
 CC studying in vivo signal transduction mechanism and identifying targets  
 CC for drug development e.g. based on olfactory and gustatory receptors in  
 CC form of agonists and antagonists by screening intrinsic and extrinsic  
 CC ligands as bitter taste inhibitors, taste enhancers and fragrance  
 CC improvers. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 235 AA;

Query Match 16.9%; Score 77; DB 5; Length 235;  
 Best Local Similarity 37.2%; Pred. No. 2.5;  
 Matches 16; Conservative 3; Mismatches 18; Indels 6; Gaps 1;

QY 19 WTGLWTGLGEGGIGPEGOASPTPDCASRWPRSRASRWPSA 61  
 DB 11 WLGLWVGL-----GLRPTFRVCSPLCGPLWPRASASLCVWGS 47

RESULT 8  
 ID ABG14023 standard; protein; 206 AA.  
 AC ABG14023;  
 XX  
 DT 18-FEB-2002 (first entry)  
 DE Novel human diagnostic protein #14014.  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US008631.  
 XX  
 PR 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS78210.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX  
 PS Claim 20; SEQ ID NO 44382; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders

CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 206 AA;

Query Match 16.8%; Score 76.5; DB 4; Length 206;  
 Best Local Similarity 33.3%; Pred. No. 2.5;  
 Matches 25; Conservative 3; Mismatches 28; Indels 19; Gaps 3;  
 QY 5 PRRVSAGC-GFADAHWTGLWTG-----LGGQGG--GIGPQGQASPTPD 45  
 DB 91 PREGXGSRVNPQRWRWAGTAGGPTPSAATGPGAKSLITRGQCGWPAFSEGFAPKPTPT 150  
 QY 46 CASRWPRGSRWPWS 60  
 DB 151 RNSSWPAASAPSGS 165

RESULT 9  
 ABG03615  
 ID ABG03615 standard; protein; 443 AA.

AC ABG03615;  
 XX  
 DT 13-FEB-2002 (first entry)  
 DE Novel human diagnostic protein #3606.

Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.  
 XX WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US008631.  
 XX  
 PR 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX  
 PA (HYSE-) HYSEQ INC.

PI Dmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS67802.

XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.

XX Claim 20; SEQ ID NO 33974; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences (II) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 443 AA;

Query Match 16.6%; Score 75.5; DB 4; Length 443;  
 Best Local Similarity 23.9%; Pred. No. 7.3;  
 Matches 26; Conservative 7; Mismatches 41; Indels 35; Gaps 4;  
 QY 5 PRRVSAGCGFADAHW-----TGLWTGLGEGGEGGIGPQGASPTPD 45  
 DB 4 PRRQSECGAPTLTWPPGSGNLPGQGASPLSASFGAGSGRGPAAAG-GSGASCTPSR 62  
 QY 46 CASRWPRGSRWP-----WSAGLT-----VRDRPQLGELCMGRG 79  
 DB 63 GPASWSRGAQVPRSSRWAGSASXNAGSPTPTTSQPPRALCAAAG 111

RESULT 10

ADC86843  
 ID ADC86843 standard; protein; 1063 AA.

AC ADC86843;  
 XX  
 DT 01-JAN-2004 (first entry)

DE Human GPCR protein SEQ ID NO:1296.

XX human; GPCR; guanosine triphosphate-binding protein coupled receptor;  
 KW gene therapy.

XX Homo sapiens.  
 XX EP1270724-A2.  
 XX  
 PD 02-JAN-2003.

XX 18-JUN-2002; 2002EP-00013517.

XX 18-JUN-2001; 2001JP-00246789.

XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.  
 PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.

XX Suwa M, Asai K, Akiyama Y, Aburatani H;

XX WPI; 2003-315783/31.  
 XX N-PSDB; ADC86842.

XX New polynucleotide, useful for preparing a composition for treating a  
 PT patient in need of increased or suppressed activity or expression of the  
 PT guanosine triphosphate-binding protein coupled receptor.

XX Claim 2; SEQ ID NO 1296; 28pp; English.

XX The invention relates to a novel polynucleotide encoding a guanosine  
 CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of  
 CC the invention may have a use in gene therapy. The polynucleotide and  
 CC polypeptide are useful for preparing a composition for treating a patient  
 CC in need of increased or suppressed activity or expression of the  
 CC guanosine triphosphate-binding protein coupled receptor. The protein  
 CC sequences shown in ADC85549-ADC87617 represent GPCR's of the invention.

SQ Sequence 1063 AA;

Query Match 16.6%; Score 75.5; DB 7; Length 1063;  
 Best Local Similarity 28.8%; Pred. No. 19;  
 Matches 32; Conservative 5; Mismatches 35; Indels 39; Gaps 8;

QY 2 CGRRP-----RVSAAGCGFADAHWTGLTGL-GEQGGIGFEGQASPTP-----DC-- 46  
 DB 164 CSVPREPCGACRVIDGCG-SDA-----GFGMPGTAASGVCGPHGRCVSPGGNFSIC 217

QY 47 -----ASRWPERSARW-----PWSAGLTVDRPQ-LGELQWGRG 79  
 DB 218 DSGFTGTTCHESEWPRTAGWGWAGLRPWLTPLASADIDDCLGQPCRNGG 268

RESULT 11

ADC86885  
 ID ADC86885 standard; protein; 1356 AA.  
 AC ADC86885;  
 DT 01-JAN-2004 (first entry)  
 DE Human GPCR protein SEQ ID NO:1338.  
 XX human; GPCR; guanosine triphosphate-binding protein coupled receptor;  
 XX Gene therapy.  
 XX Homo sapiens.  
 XX EPI270724-A2.  
 XX PD 02-JAN-2003.  
 XX PF 18-JUN-2002; 2002EP-00013517.  
 XX PR 18-JUN-2001; 2001JP-00246789.  
 XX (NAAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.  
 XX (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.  
 XX Suwa M, Asai K, Akiyama Y, Aburatani H;  
 XX WPI; 2003-315783/31.  
 XX DR N-P5DB; ADC86884.  
 XX New polynucleotide, useful for preparing a composition for treating a  
 PT patient in need of increased or suppressed activity or expression of the  
 PT guanosine triphosphate-binding protein coupled receptor.  
 XX Claim 2; SEQ ID NO 1338; 289p; English.

The invention relates to a novel polynucleotide encoding a guanosine  
 triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of  
 the invention may have a use in gene therapy. The polynucleotide and  
 polypeptide are useful for preparing a composition for treating a patient  
 in need of increased or suppressed activity or expression of the  
 CC guanosine triphosphate-binding protein coupled receptor. The protein  
 CC sequences shown in ADC85549-ADC87617 represent GPCR's of the invention.

SQ Sequence 1356 AA;

Query Match 16.6%; Score 75.5; DB 7; Length 1356;  
 Best Local Similarity 28.8%; Pred. No. 24;  
 Matches 32; Conservative 5; Mismatches 35; Indels 39; Gaps 8;

QY 2 CGRRP-----RVSAAGCGFADAHWTGLTGL-GEQGGIGFEGQASPTP-----DC-- 46  
 DB 164 CSVPREPCGACRVIDGCG-SDA-----GFGMPGTAASGVCGPHGRCVSPGGNFSIC 217

QY 47 -----ASRWPERSARW-----PWSAGLTVDRPQ-LGELQWGRG 79  
 DB 218 DSGFTGTTCHESEWPRTAGWGWAGLRPWLTPLASADIDDCLGQPCRNGG 268

RESULT 12

ABM65289  
 ID ABM65289 standard; protein; 104 AA.  
 XX ABM65289;  
 DT 20-OCT-2003 (first entry)  
 DE Propionibacterium acnes immunogenic polypeptide #29965.  
 XX Acne vulgaris; antisborrheic; dermatological; antibacterial;  
 XX immunostimulant; immune response; vaccine; immunogenic.  
 XX Propionibacterium acnes.  
 XX WO2003033515-A1.  
 XX PD 24-APR-2003.  
 XX PF 11-OCT-2002; 2002WO-US032727.  
 XX PR 15-OCT-2001; 2001US-00978825.  
 XX (CORI-) CORIXA CORP.  
 XX Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
 XX Zhang Y, Wang S, Jen S, Lodes WJ, Benson DR, Jones R, Carter D;  
 XX Barth B, Vallieve-Douglass J;  
 XX WPI; 2003-381789/36.  
 XX New Propionibacterium acnes polypeptides and polynucleotides encoding the  
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
 PT or for stimulating an immune response specific for a P. acnes protein.  
 XX Claim 7; SEQ ID NO 29965; 1481pp; English.

The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
 encoding a Propionibacterium acnes protein. The invention also relates to  
 polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
 immunogenic fragments of P. acnes polypeptides. The invention  
 additionally encompasses expression vectors and host cells comprising a  
 polynucleotide of the invention; antibodies against polypeptides of the  
 invention; fusion proteins comprising a polypeptide of the invention; a  
 method for stimulating an immune response specific for a P. acnes  
 polypeptide and an isolated T cell population comprising T cells prepared  
 via this method; a vaccine composition (comprising P. acnes polypeptides,  
 polynucleotides, antibodies, fusion proteins, T cell populations, or  
 antigen-presenting cells that express the polypeptide); a method and kit  
 for detecting or determining the presence or absence of P. acnes in a  
 patient; and a method for inhibiting the development of P. acnes in a  
 patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
 proteins, T cell populations or antigen-presenting cells that express the  
 polypeptides are useful for diagnosing, preventing or treating acne  
 vulgaris, or for stimulating an immune response specific for a P. acnes  
 protein. The polynucleotides can also be used as probes or primers for  
 nucleic acid hybridisation. The vaccine composition is useful for the  
 stimulation of an immune response against P. acnes, or for treating acne,  
 and the kit is useful for performing a diagnostic assay. The present  
 sequence represents a specifically claimed P. acnes polypeptide which is  
 thought to contain an immunogenic region. Note: The sequence data for  
 this patent did not form part of the printed specification, but was  
 obtained in electronic format directly from WIPO at  
 cc ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 104 AA;

Query Match 16.3%; Score 74.5; DB 6; Length 104;  
 Best Local Similarity 32.1%; Pred. No. 2;  
 Matches 27; Conservative 7; Mismatches 35; Indels 15; Gaps 5;

QY 3 GPRPRVACCGPADAHWTGLTGLGEGGQGGIGPBGQASPTPDCAASRWPRASRWPSAG 62  
 Db 14 GR-RBPACGCSGHPRYPEDPMPGLWRROE-AVAPEGHR-----CSSGFDSCTGSGSVAAP 67  
 QY 63 LTVDR-POLGE-----LCMG 77  
 Db 68 FTVRSRVPTLSAPLRWALLCVG 91

RESULT 13  
 ABP69309  
 ID ABP69309 standard; protein; 156 AA.  
 XX AC ABP69309;  
 XX DT 20-JAN-2003 (first entry)  
 XX DE Human polypeptide SEQ ID NO 1356.  
 XX KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
 KW cell-proliferative disorder; neurodegenerative disease; bacterial;  
 KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
 KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
 KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;  
 KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
 KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
 KW antiarthritic.  
 XX OS Homo sapiens.  
 XX PN WO200270539-A2.  
 XX PD 12-SEP-2002.  
 XX PF 05-MAR-2002; 2002WO-US005095.  
 XX PR 05-MAR-2001; 2001US-00799451.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
 PI Wehrman T, Wang J, Wang D, Drmanac RT;  
 XX WPI; 2002-759812/82.  
 XX N-PSDB; ABZ11526.  
 XX PT New polynucleotides comprising sequences assembled from expressed  
 PT sequence tags (ESTs), useful for treating cell-proliferative,  
 PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet  
 PT or coagulation disorders.  
 XX PS Claim 9; SEQ ID NO 1356; 1012pp + Sequence Listing; English.  
 XX CC The invention relates to an isolated polynucleotide (I) comprising a  
 CC nucleotide sequence selected from any of 948 sequences (ABZ1119-  
 CC ABZ12066) or their mature protein coding portion, active domain coding  
 CC protein or complementary sequences. The polynucleotides are useful for  
 CC identifying expressed genes or for physical mapping of human genome. The  
 CC encoded polypeptides (ABP68902-ABP69849) are useful as molecular weight  
 CC markers, as a food supplement, for generating antibodies, in medical  
 CC imaging, screening and diagnostic assays and for treating cell-  
 CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's  
 CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,  
 CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,  
 CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver  
 CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),  
 CC arthritis, etc. Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 156 AA;

Query Match 16.3%; Score 74.5; DB 5; Length 156;  
 Best Local Similarity 26.4%; Pred. No. 3;  
 Matches 23; Conservative 7; Mismatches 20; Indels 37; Gaps 3;  
 QY 10 AGCFADAHWTG-----LWTGL-----GEGQGG 33  
 Db 13 AGCGFLGVYVGVASCLREHAFVLVATHTHYGASAGALTATLVGCLGERGROAG 72  
 QY 34 IGPEGQASPTPDCAASRWPRASRW-PW 59  
 Db 73 GWREGFAGDGLRERRGPAASRWVW 99

RESULT 14  
 ABG14843  
 ID ABG14843 standard; protein; 180 AA.  
 XX AC ABG14843;  
 XX DT 18-FEB-2002 (first entry)  
 XX DE Novel human diagnostic protein #14834.  
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX OS Homo sapiens.  
 XX PN WO200175067-A2.  
 XX PD 11-OCT-2001.  
 XX PF 30-MAR-2001; 2001WO-US008631.  
 XX PR 31-MAR-2000; 2000US-00540217.  
 XX PR 23-AUG-2000; 2000US-00649167.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 XX N-PSDB; AAS79030.  
 XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX PS Claim 20; SEQ ID NO 45202; 103pp; English.  
 XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX



```
SQ Sequence 180 AA;
Query Match 16.3%; Score 74.5; DB 4; Length 180;
Best Local Similarity 26.3%; Pred. No. 3.5;
Matches 31; Conservative 6; Mismatches 40; Indels 41; Gaps 6;
QY 2 CGRRPRVS-AGC-----GFADAHWTGLWTG-----LGEQGGGIG 35
DB 44 CWRTRASCAPCAGATAPRAARTRSSGWCSPNSWSPASCPTTCCWCAGTGRPAATS 103
QY 36 PEGQASPT--PDC-----ASRWPRAS-----RWPASAGLTVDRPQLGELCMGR 78
DB 104 PRAFSTPTTSPCSPASTASPTPCSTASSARPPGTGTWPSAGPWPSPAPGPGR 161
```

RESULT 15

```
ABG12827
ID ABG12827 standard; protein; 180 AA.
XX AC ABG12827;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #12818.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX FN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS77014.
```

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

Claim 20; SEQ ID NO 43186; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: the sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at

ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 180 AA;

Query Match 16.3%; Score 74.5; DB 4; Length 180;  
Best Local Similarity 26.3%; Pred. No. 3.5;  
Matches 31; Conservative 6; Mismatches 40; Indels 41; Gaps 6;

QY 2 CGRRPRVS-AGC-----GFADAHWTGLWTG-----LGEQGGGIG 35  
DB 44 CWRTRASCAPCAGATAPRAARTRSSGWCSPNSWSPASCPTTCCWCAGTGRPAATS 103  
QY 36 PEGQASPT--PDC-----ASRWPRAS-----RWPASAGLTVDRPQLGELCMGR 78  
DB 104 PRAFSTPTTSPCSPASTASPTPCSTASSARPPGTGTWPSAGPWPSPAPGPGR 161

Search completed: August 13, 2004, 09:06:09  
Job time : 56 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2004, 09:05:07 ; Search time 19 Seconds  
(without alignments)  
214.655 Million cell updates/sec

Title: US-09-972-032-2

Perfect score: 456

Sequence: 1 MCGRRPRVSAGCGFADAHWT.....SAGLTVDRPQLGELCMGRG 79

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /cgm2\_6/ptodata/2/iaa/5A.COMB.pep.\*
- 2: /cgm2\_6/ptodata/2/iaa/5B.COMB.pep.\*
- 3: /cgm2\_6/ptodata/2/iaa/6A.COMB.pep.\*
- 4: /cgm2\_6/ptodata/2/iaa/6B.COMB.pep.\*
- 5: /cgm2\_6/ptodata/2/iaa/PCTUS.COMB.pep.\*
- 6: /cgm2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	86.5	19.0	326	4	US-09-252-991A-21802
2	78.5	17.2	451	4	US-09-252-991A-27602
3	75.5	16.6	423	4	US-09-252-991A-24347
4	75.5	16.6	663	4	US-09-252-991A-30843
5	75	16.4	431	4	US-09-252-991A-26751
6	74.5	16.3	312	4	US-09-252-991A-31204
7	73.5	16.1	170	4	US-09-193-637A-299
8	73	16.0	138	4	US-09-252-991A-24447
9	73	16.0	436	4	US-09-252-991A-33064
10	72.5	15.9	148	4	US-09-252-991A-17048
11	72.5	15.9	719	4	US-09-252-991A-30052
12	71	15.6	330	4	US-09-252-991A-25986
13	71	15.6	860	4	US-09-252-991A-25911
14	70.5	15.5	141	4	US-09-252-991A-16951
15	70	15.4	264	4	US-09-252-991A-24262
16	70	15.4	532	1	US-08-494-168-9
17	69	15.1	153	4	US-09-252-991A-21201
18	69	15.1	178	4	US-09-252-991A-23713
19	69	15.1	1217	4	US-09-252-991A-26104
20	68	14.9	169	4	US-09-252-991A-25693
21	68	14.9	169	4	US-09-252-991A-25543
22	67.5	14.8	121	4	US-09-072-596-253
23	67.5	14.8	121	4	US-09-072-967-258
24	67.5	14.8	203	4	US-09-252-991A-21364
25	67.5	14.8	239	4	US-09-252-991A-28985
26	67.5	14.8	676	4	US-09-252-991A-23181
27	67.5	14.8	1958	1	US-07-945-283-2

28	67	14.7	223	4	US-09-252-991A-20332	Sequence 20332, A
29	67	14.7	372	4	US-09-252-991A-20108	Sequence 20108, A
30	67	14.7	475	4	US-09-252-991A-22247	Sequence 22247, A
31	67	14.7	728	4	US-09-252-991A-23613	Sequence 23613, A
32	66.5	14.6	117	4	US-09-252-991A-20556	Sequence 20556, A
33	66.5	14.6	196	4	US-09-252-991A-19503	Sequence 19503, A
34	66.5	14.6	401	4	US-09-252-991A-21672	Sequence 21672, A
35	66	14.5	192	4	US-09-252-991A-27287	Sequence 27287, A
36	66	14.5	206	4	US-09-252-991A-31481	Sequence 31481, A
37	66	14.5	274	4	US-09-252-991A-23172	Sequence 23172, A
38	66	14.5	371	4	US-09-252-991A-29793	Sequence 29793, A
39	66	14.5	518	4	US-09-252-991A-23604	Sequence 23604, A
40	66	14.5	533	4	US-09-252-991A-23560	Sequence 23560, A
41	65.5	14.4	112	4	US-09-489-039A-7360	Sequence 7360, Ap
42	65.5	14.4	125	4	US-09-252-991A-27936	Sequence 27936, A
43	65.5	14.4	133	4	US-09-252-991A-27120	Sequence 27120, A
44	65.5	14.4	193	4	US-09-252-991A-28920	Sequence 28920, A
45	65.5	14.4	389	2	US-08-485-449-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1

US-09-252-991A-21802  
; Sequence 21802, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 21802  
; LENGTH: 326  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-21802

Query Match 19.0%; Score 86.5; DB 4; Length 326;  
Best Local Similarity 42.1%; Pred. No. 0.055;  
Matches 24; Conservative 3; Mismatches 27; Indels 3; Gaps 2;  
QY 4 RPRRVAG--CGFADAHWTGLTGLGEGGEGGIGPEGQASPTPDCASRWPRASRW 58  
DB 261 RNRPRAGAACAPAI PHSITRIATGSSRWAGKAAPV-RAGRRPGGASRWPRPGRWP 316

RESULT 2

US-09-252-991A-27602  
; Sequence 27602, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 27602  
; LENGTH: 451  
; TYPE: PRT

```

; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27602

Query Match
Best Local Similarity 17.2%; Score 78.5; DB 4; Length 451;
Matches 28; Conservative 3; Mismatches 38; Indels 13; Gaps 3;

QY 5 PRRVSAGCGFADAHWTGLGEGGEGIGPEGQASPTTDCASRWPRSRWPSAGLT 64
DB 142 PRLAFGAAGROPFAH-----GAGEKQFADRLFAAQRTPRPACAGPATGAEPRPGFGRP 195

QY 65 VR-----DRPO--LGELCMGRG 79
DB 196 VRRPAGAEPRQRPAGQAPFGRG 217

RESULT 3
US-09-252-991A-24347
; Sequence 24347, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24347
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24347

Query Match
Best Local Similarity 16.6%; Score 75.5; DB 4; Length 423;
Matches 19; Conservative 6; Mismatches 14; Indels 15; Gaps 2;

QY 40 ASPTDCASRWPRSRASRW------WSAGLTVRDQPGELCMGRG 79
DB 3 ASPSSVWASRW-RSPTRWPTSSRTTTTPAAFSNWTAAACACRPRPRKEAPLSRG 55

RESULT 4
US-09-252-991A-30843
; Sequence 30843, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30843
; LENGTH: 663
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30843

Query Match
Best Local Similarity 16.6%; Score 75.5; DB 4; Length 663;
Matches 15; Conservative 1; Mismatches 10; Indels 1; Gaps 1;

QY 35 GPEGQASPTTDCASRWPRSRASRWPSA 61
DB 110 GP-SPASPAPAAACRRWRPRAACHWPASA 135

RESULT 5
US-09-252-991A-26751
; Sequence 26751, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26751
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26751

Query Match
Best Local Similarity 16.4%; Score 75; DB 4; Length 431;
Matches 27; Conservative 6; Mismatches 27; Indels 30; Gaps 5;

QY 2 CGPRRVSGCGCFADA-----HWTGLWTGLGEGGEGIGPEGQASPTTDC-ASRW 51
DB 12 CDRPCTAGTPTSWRTAPDRAPARWWRGSPGPRRGRHHPWPE-----PGCRSRRP 65

QY 52 RSAS-----RWP-----WSAGLTVRD 67
DB 66 RSALAPGEADRAPWRWPDACRWTAPSLRD 95

RESULT 6
US-09-252-991A-31204
; Sequence 31204, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31204
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31204

Query Match
Best Local Similarity 16.3%; Score 74.5; DB 4; Length 312;
Matches 22; Conservative 2; Mismatches 21; Indels 13; Gaps 2;

QY 4 RPRRVSGCGCFADAHWTGLGEGGEGIGPEGQASPTTDCASRWPRSRASRWPSA 61
DB 210 RPRASHRGIS-----GRGAGSGPADVCPAPDAAGR--RGAIRSPGGA 254

RESULT 7
US-09-199-637A-299

```

; Sequence 299, Application US/09199637A  
; Patent No. 655411  
; GENERAL INFORMATION:  
; APPLICANT: Ausubel, Frederick  
; APPLICANT: Goodman, Howard M.  
; APPLICANT: Rahme, Laurence G.  
; APPLICANT: Mahajan-Miklos, Shalina  
; APPLICANT: Tan, Man-Wah  
; APPLICANT: Cao, Hui  
; APPLICANT: Drenkard, Eliana  
; APPLICANT: Tsorgalis, John  
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF  
; FILE REFERENCE: 00786/361002  
; CURRENT APPLICATION NUMBER: US/09/199,637A  
; CURRENT FILING DATE: 1998-11-25  
; PRIOR APPLICATION NUMBER: 60/066,517  
; PRIOR FILING DATE: 1997-11-25  
; NUMBER OF SEQ ID NOS: 437  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 299  
; LENGTH: 170  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-199-637A-299

Query Match 16.1%; Score 73.5; DB 4; Length 170;  
Best Local Similarity 32.1%; Pred. No. 0.75;  
Matches 18; Conservative 5; Mismatches 20; Indels 13; Gaps 1;  
  
QY 3 GRPRVAGCGFADAHWTGLTGLGEGGIGGEGQASPTPDCASRWPSASRWP 58  
DB 16 GWPRMPAGAGWRACWRRWT-----AWACPATNCWCGRRRSARRP 58

RESULT 8  
US-09-252-991A-24447  
; Sequence 24447, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 24447  
; LENGTH: 138  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-24447

Query Match 16.0%; Score 73; DB 4; Length 138;  
Best Local Similarity 32.4%; Pred. No. 0.68;  
Matches 23; Conservative 8; Mismatches 30; Indels 10; Gaps 4;  
  
QY 6 RVSAGCGFADAHWTG--LWTGLGEGQ-----EGGIGGEGQAS--PTPDCAS--RWPSAS 55  
DB 31 RSAAPGSPASAPWPGAAVWRRTGRTQRSAGRSRFSRGAAGRSAPRGRSRPQSGVA 90  
  
QY 56 RWPWSAGLTVR 66  
DB 91 GWPGARYSSR 101

RESULT 9  
US-09-252-991A-33064  
; Sequence 33064, Application US/09252991A

; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 33064  
; LENGTH: 436  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-33064

Query Match 16.0%; Score 73; DB 4; Length 436;  
Best Local Similarity 33.3%; Pred. No. 2.4;  
Matches 24; Conservative 6; Mismatches 20; Indels 22; Gaps 3;  
  
QY 6 RVSAGCGFADAHWTGLTGLGEGGIGGEGQASPTPDCASRWPSASRWPWSA--- 61  
DB 255 RVAA-----QLPVGLEAGQGQGGPLQRAEPGTDAASQ-----PWQAPGV 296  
  
QY 62 GLTVDRDPOLGE 73  
DB 297 GVLEQEAUVGE 308

RESULT 10  
US-09-252-991A-17048  
; Sequence 17048, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 17048  
; LENGTH: 148  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17048

Query Match 15.9%; Score 72.5; DB 4; Length 148;  
Best Local Similarity 31.2%; Pred. No. 0.83;  
Matches 20; Conservative 3; Mismatches 30; Indels 11; Gaps 3;  
  
QY 2 CGPRPRVSAGC-----GFADAHWTGLTGLGEGGIGGEGQASPTPDCASRWP---RS 53  
DB 62 CGTPRATPCNCTAAAGNTSASWSRSARMSASNSASSP---ASPTPCVSTWGXACRA 118  
  
QY 54 ASRW 57  
DB 119 ASGW 122

RESULT 11  
US-09-252-991A-30052  
; Sequence 30052, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 30052

; LENGTH: 719

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-30052

Query Match 15.9%; Score 72.5; DB 4; Length 719;  
 Best Local Similarity 26.7%; Pred. No. 4.8;  
 Matches 28; Conservative 10; Mismatches 38; Indels 29; Gaps 5;

QY 3 GRPRVSAGCGFADAHWTGLTGLGEGEGGI-----GPEGQASPT-----PDCASR 49

DB 11 GPRRRRGCGCRACRAVPVA-WSGVGSAPRGVLRPPPAALRRGGDDRQAATGRRAPGCAAR 69

QY 50 WPSASR-----WPW-----SAGLTVDRDPQGLGELCMGRG 79

DB 70 WRPAAGRAIAATGAAVPGVAPRGAAAGVPAASDRRRARAPGAG 114

# RESULT 12

US-09-252-991A-25986

; Sequence 25986, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 25986

; LENGTH: 330

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-25986

Query Match 15.6%; Score 71; DB 4; Length 330;  
 Best Local Similarity 29.7%; Pred. No. 3;  
 Matches 30; Conservative 3; Mismatches 22; Indels 46; Gaps 5;

QY 4 RPRVSAGCGFADAHWTGLTGLGEGEG-----GIGPEGQAS-----PT 43

DB 228 RPRPGG-----OPGLRGQEGPGRSRQGRPDRAPEGARRAGQSRPRFP 276

QY 44 PDCASRWPSASRPW-----SAGL--TVDRP 69

DB 277 PDRFAGDPHRRRTGPWADAPDRGQRAESAGLEADLRDQP 317

# RESULT 13

US-09-252-991A-25911

; Sequence 25911, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 25911

; LENGTH: 860

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-25911

Query Match 15.6%; Score 71; DB 4; Length 860;  
 Best Local Similarity 40.5%; Pred. No. 8.6;  
 Matches 17; Conservative 3; Mismatches 18; Indels 4; Gaps 1;

QY 36 PEGQASPTPDCASRWPSASRWPSAGLTVDRDPQLGELCMG 77

DB 781 PPSVSPTFASMTWRWTATRWYSSAD-----SRPQITILAE 818

# RESULT 14

US-09-252-991A-16951

; Sequence 16951, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 16951

; LENGTH: 141

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-16951

Query Match 15.5%; Score 70.5; DB 4; Length 141;  
 Best Local Similarity 28.4%; Pred. No. 1.3;  
 Matches 19; Conservative 5; Mismatches 18; Indels 25; Gaps 3;

QY 36 PEGQASPTPDCAS-----RWP-----RSASRWPSAGLTVDRPQ--- 70

DB 27 PSGRGAPARGASWMPGPGFQWPGPSRSGAPRRHAGRGARWFWPCGCTARGWPRSPC 86

QY 71 LGELCMG 77

DB 87 LPRRCPG 93

# RESULT 15

US-09-252-991A-24262

; Sequence 24262, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

```
; SEQ ID NO 24262
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24262

Query Match      15.4% Score 70; DB 4; Length 264;
Best Local Similarity 35.5%; Pred. No. 3;
Matches 22; Conservative 5; Mismatches 25; Indels 10; Gaps 3;

QY      5 PRRYSAGCGFADAHWTGLWTGLGEGGIGPEGQASPTPDCASRWPRSASRWPSAGLT 64
Db      100 PLRAGAVAGSA-----ARLGAADQPLRG-KGRDRWFCCARGWP-TASCWTWSAGRS 149

QY      65 VR 66
Db      150 SR 151
```

Search completed: August 13, 2004, 09:08:19  
Job time : 20 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2004, 09:06:12 ; Search time 46 Seconds  
(without alignments)  
539.136 Million cell updates/sec

Title: US-09-972-032-2

Perfect score: 456

Sequence: 1 MCGPRRVSGCGFADAHWT.....SAGLTVDRLQGLCMGRG 79

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1292805 seqs, 313927144 residues

Total number of hits satisfying chosen parameters: 1292805

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

```
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	100.0	79	12	US-09-972-032-2
2	79.5	17.4	123	16	US-10-437-963-125319
3	79	17.3	99	16	US-10-437-963-175071
4	79	17.3	813	16	US-10-437-963-182704
5	78	17.1	11088	16	US-10-203-295-7
6	77	16.9	215	15	US-10-094-749-1759
7	77	16.9	235	12	US-10-343-650A-98
8	75.5	16.6	146	16	US-10-437-963-190988
9	75.5	16.6	215	12	US-10-425-114-70579
10	75.5	16.6	231	16	US-10-437-963-134598
11	75.5	16.6	1063	14	US-10-017-161-1624
12	75.5	16.6	1063	15	US-10-292-798-1296
13	75.5	16.6	1356	14	US-10-017-161-1678
14	75.5	16.6	1356	15	US-10-292-798-1338
15	73.5	16.1	170	10	US-09-975-719-299

16	73	16.0	515	16	US-10-437-963-187892	Sequence 187892,
17	72	15.8	514	12	US-10-296-115-1432	Sequence 1432, Ap
18	71	15.6	253	11	US-09-833-245-1132	Sequence 1132, Ap
19	71	15.6	267	14	US-10-156-761-10080	Sequence 10080, A
20	71	15.6	625	14	US-10-156-761-15008	Sequence 15008, A
21	71	15.6	1081	16	US-10-437-963-196915	Sequence 196915,
22	70	15.4	177	12	US-10-425-114-42696	Sequence 42696, A
23	70	15.4	182	12	US-10-425-114-51184	Sequence 51184, A
24	70	15.4	222	12	US-10-425-114-64974	Sequence 64974, A
25	70	15.4	569	14	US-10-156-761-14513	Sequence 14513, A
26	70	15.4	677	14	US-10-259-165-192	Sequence 192, App
27	70	15.4	1214	12	US-10-267-502-333	Sequence 333, App
28	70	15.4	1669	12	US-09-918-715-252	Sequence 252, App
29	70	15.4	1869	15	US-10-372-683-8	Sequence 8, Appli
30	69.5	15.2	552	14	US-10-106-698-6365	Sequence 6365, Ap
31	69	15.1	122	16	US-10-437-963-127558	Sequence 127558,
32	69	15.1	141	16	US-10-437-963-127488	Sequence 127488,
33	69	15.1	190	12	US-10-425-114-72793	Sequence 72793, A
34	68.5	15.0	125	16	US-10-437-963-186225	Sequence 186225,
35	68.5	15.0	179	16	US-10-437-963-176622	Sequence 176622,
36	68.5	15.0	207	15	US-10-104-047-2751	Sequence 2751, Ap
37	68.5	15.0	207	16	US-10-471-115-7	Sequence 7, Appli
38	68.5	15.0	236	16	US-10-437-963-178641	Sequence 178641,
39	68.5	15.0	2169	9	US-09-738-626-5455	Sequence 5455, Ap
40	68	14.9	176	15	US-10-108-260A-2704	Sequence 2704, Ap
41	68	14.9	280	12	US-10-425-114-45512	Sequence 45512, A
42	68	14.9	304	12	US-10-399-645-5	Sequence 5, Appli
43	68	14.9	316	12	US-10-262-511-118	Sequence 118, App
44	67.5	14.8	121	14	US-10-193-002-253	Sequence 253, App
45	67.5	14.8	121	14	US-10-084-843-258	Sequence 258, App

#### ALIGNMENTS

#### RESULT 1

```
US-09-972-032-2
; Sequence 2, Application US/09972032
; Publication No. US20020086361A1
; GENERAL INFORMATION:
; APPLICANT: Case Western Reserve University
; APPLICANT: Montano, Monica
; APPLICANT: Sutton, Amelia
; TITLE OF INVENTION: A Modulator of Antiestrogen Pharmacology
; FILE REFERENCE: 27708/04003
; CURRENT APPLICATION NUMBER: US/09/972,032
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 60/238,190
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 2
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-972-032-2
```

Query Match	100.0%	Score	456;	DB	12;	Length	79;
Best Local Similarity	100.0%	Pred. No.	1.7e-37;				
Matches	79;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MCGPRRVSGCGFADAHWTGLTGLGEGQBGIGEGQASPTDPCASRWPSASRWPS	60				
Db	1	MCGPRRVSGCGFADAHWTGLTGLGEGQBGIGEGQASPTDPCASRWPSASRWPS	60				
Qy	61	AGLTVDRLQGLCMGRG	79				
Db	61	AGLTVDRLQGLCMGRG	79				
RESULT 2							
US-10-437-963-125319							
; Sequence 125319, Application US/10437963							



```
; SEQ ID NO 7
; LENGTH: 11088
; TYPE: PRT
; ORGANISM: Streptomyces noursei ATCC 11455
US-10-203-295-7

Query Match      17.1%; Score 78; DB 16; Length 11088;
Best Local Similarity 37.2%; Pred. No. 2.9e+02;
Matches 32; Conservative 7; Mismatches 33; Indels 14; Gaps 6;

QY 4 RPRVSGCGFADAHWTGL---WTGLGEGGEGGIGPEGQASPT-PDCASR-WPRS--ASR 56
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 828 RPELSAVTGARAHVRGVTWRAGLFD----GTGARRADLTYFFQHQRPWPTAARAAQ 883

QY 57 WPGSAGLTVDRPQLG---ELCMGRG 79
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 884 DVTAAGLGAADHPLLGATVELADGAG 909

RESULT 6
US-10-094-749-1759
; Sequence 1759, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUOI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUKIO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAGURU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOKUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1759
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-1759

Query Match      16.9%; Score 77; DB 15; Length 215;
Best Local Similarity 31.9%; Pred. No. 8.4;
Matches 23; Conservative 3; Mismatches 28; Indels 18; Gaps 4;

QY 7 RVSGCGFADAHWTGL-----WTGLGEGGEGGIGPEGQASPTPD-CASRWPRSASRW 57
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 64 RCGARC-----WEGSSLAQLQPPWTPSGSLTAAGVQVAD---SCWRSWERSCFRW 114

QY 58 PWSAGLTVDRRP 69
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 115 PPSSTAARSSP 126

RESULT 7
US-10-343-650A-98

; Sequence 98, Application US/10343650A
; Publication No. US20040067499A1
; GENERAL INFORMATION:
; APPLICANT: HAGA, TATSUYA
; TITLE OF INVENTION: NOVEL G-PROTEIN COUPLED RECEPTOR
; FILE REFERENCE: 31671-186347 US/10/343,650A
; CURRENT APPLICATION NUMBER: 2003-07-21
; PRIOR APPLICATION NUMBER: JP 2000/237818
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: JP 2001/34434
; PRIOR FILING DATE: 2001-02-13
; NUMBER OF SEQ ID NOS: 694
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 98
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-343-650A-98

Query Match      16.9%; Score 77; DB 12; Length 235;
Best Local Similarity 37.2%; Pred. No. 9.2;
Matches 16; Conservative 3; Mismatches 18; Indels 6; Gaps 1;

QY 19 WTGLWTGLGEGGIGPEGQASPTPD-CASRWPRSASRWPSA 61
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 11 WLGLWVGL-----GLRPTFRVCSPELCLPLWRPSASLCVWGS 47

RESULT 8
US-10-437-963-190988
; Sequence 190988, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 190988
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(146)
; OTHER INFORMATION: unsure at all xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_8734C.1.pep
US-10-437-963-190988

Query Match      16.6%; Score 75.5; DB 16; Length 146;
Best Local Similarity 31.9%; Pred. No. 8.2;
Matches 22; Conservative 5; Mismatches 19; Indels 23; Gaps 3;

QY 3 GRPRRYVSGCGFADAHWTGLWTGLGEGGEGGIGPEGQASPTP----DCASRWPR----- 52
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 9 GRPRRAGXXYG-----GRGQREGVLPQGRAPRPHGRASPVARRRRLPLV 57

QY 53 --SASRWPW 59
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 58 PLSLESSPW 66
```

```
RESULT 9
US-10-425-114-70579
; Sequence 70579, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 70579
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3607-033-A2_FLI.pep
US-10-425-114-70579

Query Match 16.6%; Score 75.5; DB 12; Length 215;
Best Local Similarity 35.7%; Pred. No. 12; Mismatches 39; Indels 11; Gaps 3;
Matches 30; Conservative 4;

QY 3 GRPRVSAGCGFADAHWTGLTGLGEGGIGPEGO--ASPTPDCASRWPRASRWPS 60
DB 76 GGEQAVRAGRA-----RPLPPLGEGGAGHPGDPGARGQPLQEPGRRPADRRPGAH 131
QY 61 AGLTVRDR-----POLGELCMGRG 79
DB 132 AGLRRGRGRGLVHPAVGRLLRPG 155

RESULT 10
US-10-437-963-134598
; Sequence 134598, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 134598
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_36356C.1.pep
US-10-437-963-134598

Query Match 16.6%; Score 75.5; DB 16; Length 231;
Best Local Similarity 25.8%; Pred. No. 13; Mismatches 21; Indels 45; Gaps 3;
Matches 24; Conservative 3;

QY 4 RPRVSAGCGFADAHWTGLTGLGEGGIGGIP-----EGQASPTPD 45
DB 41 RRRLWAAAAADRRWT-----EGDRGGIGLPHGDEVTISLPIRLAVAAAGPAPA 93
```

```
QY 46 CASRWP-----RSASRWP 58
DB 94 RRRRWPPLPLSPALGRPPADGCRPVPPTARRWP 126

RESULT 11
US-10-017-161-1624
; Sequence 1624, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1624
; LENGTH: 1063
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-161-1624

Query Match 16.6%; Score 75.5; DB 14; Length 1063;
Best Local Similarity 28.8%; Pred. No. 54; Mismatches 32; Conservative 5; Indels 39; Gaps 8;
Matches 32;

QY 2 CCRPR-----RVSAGCGFADAHWTGLTGL-CEGEGGIGPEGOASPTP-----DC-- 46
DB 164 CSVPREPCGGACRVIDCGG-SDA-----GPGMPGTAAAGVCGPHGRCVSPQGNFSCIC 217
QY 47 -----ASRWPRASRW-----PWSAGLTVRDRPQ-LGELCMGRG 79
DB 218 DSGFTGTTCHESENPRTAGWWGAGLRPLTPLASADIDDCLGCPCRNGG 268

RESULT 12
US-10-292-798-1296
; Sequence 1296, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1296
; LENGTH: 1063
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-292-798-1296

Query Match 16.6%; Score 75.5; DB 15; Length 1063;
Best Local Similarity 28.8%; Pred. No. 54; Mismatches 32; Conservative 5; Indels 39; Gaps 8;
Matches 32;

QY 2 CCRPR-----RVSAGCGFADAHWTGLTGL-CEGEGGIGPEGOASPTP-----DC-- 46
DB 164 CSVPREPCGGACRVIDCGG-SDA-----GPGMPGTAAAGVCGPHGRCVSPQGNFSCIC 217
```

QY 47 -----ASRWRPGASRW-----PWSAGLTVDRPQ-LGELCMGRG 79  
DB 218 DSGFTGYCHSEWPRTAGWWGAGLRPWLTPPLASADIDDCLGQPCRNGG 268

## RESULT 13

US-10-017-161-1678  
; Sequence 1678, Application US/10017161  
; Publication No. US20030143668A1  
; GENERAL INFORMATION:  
; APPLICANT: SUWA, MAKIKO  
; APPLICANT: ASAI, KIYOSHI  
; APPLICANT: ABURATANI, HIROYUKI  
; APPLICANT: AKIYAMA, YUTAKA  
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS  
; FILE REFERENCE: 084335/0152  
; CURRENT APPLICATION NUMBER: US/10/017,161  
; CURRENT FILING DATE: 2002-12-18  
; PRIOR APPLICATION NUMBER: JP 2001/246789  
; PRIOR FILING DATE: 2001-06-18  
; NUMBER OF SEQ ID NOS: 2430  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1678  
; LENGTH: 1356  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-017-161-1678

Query Match 16.6%; Score 75.5; DB 14; Length 1356;  
Best Local Similarity 28.8%; Pred. No. 68;  
Matches 32; Conservative 5; Mismatches 35; Indels 39; Gaps 8;  
QY 2 CGRPR-----RVSAAGCGFADAHWTGLWTGL-GEQGGIGPEGQASPTP-----DC-- 46  
DB 164 CSVPRPCGGACRVIDGCG-SDA-----GPGMPTAASGVCGPHGRCVSPGQGNFSCIC 217  
QY 47 -----ASRWRPGASRW-----PWSAGLTVDRPQ-LGELCMGRG 79  
DB 218 DSGFTGYCHSEWPRTAGWWGAGLRPWLTPPLASADIDDCLGQPCRNGG 268

## RESULT 14

US-10-292-798-1338  
; Sequence 1338, Application US/10292798  
; Publication No. US20030235833A1  
; GENERAL INFORMATION:  
; APPLICANT: SUWA, MAKIKO  
; APPLICANT: ASAI, KIYOSHI  
; APPLICANT: AKIYAMA, YUTAKA  
; APPLICANT: ABURATANI, HIROYUKI  
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS  
; FILE REFERENCE: 084335/166  
; CURRENT APPLICATION NUMBER: US/10/292,798  
; CURRENT FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: 10/017,161  
; PRIOR FILING DATE: 2001-12-18  
; PRIOR APPLICATION NUMBER: JP 2001-246789  
; PRIOR FILING DATE: 2001-06-18  
; NUMBER OF SEQ ID NOS: 2070  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1338  
; LENGTH: 1356  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-292-798-1338

Query Match 16.6%; Score 75.5; DB 15; Length 1356;  
Best Local Similarity 28.8%; Pred. No. 68;  
Matches 32; Conservative 5; Mismatches 35; Indels 39; Gaps 8;  
QY 2 CGRPR-----RVSAAGCGFADAHWTGLWTGL-GEQGGIGPEGQASPTP-----DC-- 46

DB 164 CSVPRPCGGACRVIDGCG-SDA-----GPGMPTAASGVCGPHGRCVSPGQGNFSCIC 217  
QY 47 -----ASRWRPGASRW-----PWSAGLTVDRPQ-LGELCMGRG 79  
DB 218 DSGFTGYCHSEWPRTAGWWGAGLRPWLTPPLASADIDDCLGQPCRNGG 268

## RESULT 15

US-09-975-719-299  
; Sequence 299, Application US/09975719  
; Publication No. US20030022349A1  
; GENERAL INFORMATION:  
; APPLICANT: Ausubel, Frederick M.  
; APPLICANT: Rahme, Laurence G.  
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF  
; FILE REFERENCE: 00786/361003  
; CURRENT APPLICATION NUMBER: US/09/975,719  
; CURRENT FILING DATE: 2001-10-10  
; PRIOR APPLICATION NUMBER: US 09/199,637  
; PRIOR FILING DATE: 1998-11-25  
; PRIOR APPLICATION NUMBER: US 60/066,517  
; PRIOR FILING DATE: 1997-11-25  
; NUMBER OF SEQ ID NOS: 437  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 299  
; LENGTH: 170  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-975-719-299

Query Match 16.1%; Score 73.5; DB 10; Length 170;  
Best Local Similarity 32.1%; Pred. No. 15;  
Matches 18; Conservative 5; Mismatches 20; Indels 13; Gaps 1;  
QY 3 GRPRVSAAGCGFADAHWTGLWTGLGEGGGIGPEGQASPTPDCASRWPRSRWP 58  
DB 16 GWPMPYPAGWRRACWRRWT-----AWACPATNCWSGWRGRRARP 58

Search completed: August 13, 2004, 09:09:17  
Job time : 47 secs



Result No.	Score	Query Match	Length	DB	ID	Description
1	73	16.0	1104	2	T49647	MSPI related prote
2	70.5	15.5	905	2	F82734	serine proteinase
3	70.5	15.5	910	2	H82826	serine proteinase
4	70	15.4	1669	1	CGHU4B	collagen alpha 1(I
5	68	14.9	282	2	T35294	probable endo alph
6	67.5	14.8	1733	1	B45344	probable nuclear a
7	67.5	14.8	1958	2	B40582	hypothetical prote
8	67	14.7	425	2	A70889	probable mce3 prot
9	67	14.7	762	2	A54411	amine oxidase (cop
10	67	14.7	1572	2	T00027	brain-specific ang
11	66	14.5	1003	2	T113951	DNA topoisomerase
12	66	14.5	1154	2	A53275	hypothetical prote
13	65.5	14.4	219	2	AG1940	hypothetical prote
14	65.5	14.4	268	2	T30630	hypothetical prote
15	65.5	14.4	1034	2	A24925	beta-galactosidase
16	65.5	14.4	1157	2	F97255	fusion of alpha-gl
17	64.5	14.1	118	2	S3491	Ig heavy chain - h
18	64.5	14.1	152	2	T34649	hypothetical prote
19	64.5	14.1	1669	1	CGMS4B	collagen alpha 1(I
20	63.5	13.9	114	2	T26000	hypothetical prote
21	63.5	13.9	274	2	A45754	tryptase (EC 3.4.2
22	63	13.8	191	2	D72711	hypothetical prote
23	63	13.8	307	2	S47221	catechol 2,3-dioxy
24	63	13.8	359	2	G02221	CAGRI protein - hu
25	63	13.8	376	2	S35247	fimbrial adhesin f
26	63	13.8	376	2	S42747	fimbrial adhesin f
27	63	13.8	1216	2	A55620	apical endosomal p
28	63	13.8	1838	1	CGHU1V	collagen alpha 1(V
29	63	13.8	1843	2	S18803	collagen alpha 1(V

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.  
A;Authors: Martins, E.M.F.; Matukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A;Reference number: A59328  
A;Contents: annotation  
A;Genetics:  
A;Gene: XF1026  
C;Superfamily: Autotransporter subtilisin-like protease precursor; subtilisin homology

Query Match 15.5%; Score 70.5; DB 2; Length 905;  
Best Local Similarity 27.2%; Pred. No. 15;  
Matches 22; Conservative 8; Mismatches 22; Indels 29; Gaps 5;

QY 7 RVSAACGCF-----ADAHWTGL-WTGLGEGQ-----EGGIGPEGQASP 42  
Db 786 REQSGSGFLQVQSAQATRSQLLAGVRTWEGWAGVQWGRGEGWQTLRQSLNPNQASFTA 845

QY 43 TPDCASRW-PRSASRWPNWSAG 62  
Db 846 T-----SSWTPLVASSWPGRSG 862

RESULT 3  
H82826  
serine proteinase XP0267 [imported] - Xylella fastidiosa (strain 9a5c)  
C;Species: Xylella fastidiosa  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 24-Nov-2003  
C;Accession: H82826  
R;Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
Nature 406, 151-157, 2000  
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A;Reference number: A82515; MUID:20365717; PMID:10910347  
A;Note: for a complete list of authors see reference number A59328 below  
A;Accession: H82826  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-910 <SIM>  
A;Cross-references: GB:A8003880; GB:A8003849; NID:G9105080; PIDN:AAP83080.1; GSPDB:GN001  
A;Experimental source: strain 9a5c  
R;Simpson, A.J.G.; Reinach, P.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H  
as-Neto, E.; Docena, C.; El-Dorty, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, C.Y.  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A;Reference number: A59328  
A;Contents: annotation  
A;Genetics:  
A;Gene: XF0267  
C;Superfamily: Autotransporter subtilisin-like protease precursor; subtilisin homology

Query Match 15.5%; Score 70.5; DB 2; Length 910;  
Best Local Similarity 27.2%; Pred. No. 15;  
Matches 22; Conservative 8; Mismatches 22; Indels 29; Gaps 5;

QY 7 RVSAACGCF-----ADAHWTGL-WTGLGEGQ-----EGGIGPEGQASP 42  
Db 791 REQSGSGFLQVQSAQATRSQLLAGVRTWEGWAGVQWGRGEGWQTLRQSLNPNQASFTA 850

QY 43 TPDCASRW-PRSASRWPNWSAG 62  
Db 851 T-----SSWTPLVASSWPGRSG 867

RESULT 4  
CGHU4B  
collagen alpha 1(IV) chain precursor - human  
N;Alternate names: procollagen alpha 1(IV) chain  
C;Species: Homo sapiens (man)  
C;Date: 28-May-1986 #sequence\_revision 31-Dec-1992 #text\_change 07-Dec-1999  
C;Accession: S16876; A32117; S02738; S00048; S25826; A23115; S00207; S35614; A02863; A58  
R;Soininen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.  
J. Biol. Chem. 264, 13565-13571, 1989  
A;Title: Structural organization of the gene for the alpha-1 chain of human type IV colla  
A;Reference number: S16876; MUID:89340433; PMID:2701944  
A;Accession: S16876  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-1669 <SO11>  
A;Cross-references: EMBL:J04217; GB:J05039; NID:G180800; PIDN:AA53098.1; PID:G180800  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1988  
R;Soininen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Tryggvason, K.  
J. Biol. Chem. 263, 17217-17220, 1988  
A;Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen are  
A;Reference number: A92690; MUID:89034231; PMID:3182844  
A;Accession: A32117  
A;Molecule type: DNA  
A;Residues: 1-28 <SO12>  
A;Cross-references: EMBL:J04217; NID:G180759; PIDN:AAA53097.1; PID:G553233  
R;Roesschl, E.; Pollner, R.; Kuehn, K.  
EMBO J. 7, 2687-2695, 1988  
A;Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane co  
A;Reference number: S02738; MUID:89030632; PMID:2846280  
A;Accession: S02738  
A;Status: translation not shown  
A;Molecule type: DNA  
A;Residues: 1-6, 'L', 8-28 <POE>  
A;Cross-references: EMBL:X12784; NID:G30072  
R;Brace, D.; Oberbaumer, I.; Bieringer, H.; Babel, W.; Glanville, R.W.; Deutzmann, R.;  
Eur. J. Biochem. 168, 529-536, 1987  
A;Title: Completion of the amino acid sequence of the alpha1 chain of human basement mem  
A;Reference number: S00048; MUID:88029471; PMID:3311751  
A;Accession: S00048  
A;Molecule type: mRNA  
A;Residues: 1-318, 'A', 320-944 <BRAL>  
A;Cross-references: EMBL:X05561; NID:G30066; PIDN:CAA29075.1; PID:G30067  
A;Accession: S25826  
A;Molecule type: Protein  
A;Residues: 271-318, 'A', 320-554 <BRA2>  
R;Glanville, R.W.; Qian, R.Q.; Siebold, B.; Risteli, J.; Kuehn, K.  
Eur. J. Biochem. 152, 213-219, 1985  
A;Title: Amino acid sequence of the N-terminal aggregation and cross-linking region (7S  
A;Reference number: A23115; MUID:86004708; PMID:4043082  
A;Accession: A23115  
A;Molecule type: Protein  
A;Residues: 28-236, 'KE', 239-240, 'K', 242-243 <GLA>  
A;Experimental source: placenta  
A;Note: the amino end of the mature form is blocked  
R;Soininen, R.; Haka-Riska, T.; Prockop, D.J.; Tryggvason, K.  
FEBS Lett. 225, 188-194, 1987  
A;Title: Complete primary structure of the alpha(1)-chain of human basement membrane (VII  
A;Reference number: S00207; MUID:88083584; PMID:3691802  
A;Accession: S00207  
A;Molecule type: mRNA  
A;Residues: 244-530 <SO13>  
A;Cross-references: EMBL:Y00706; NID:G29548; PIDN:CAA68698.1; PID:G29549  
R;Eble, J.A.; Golbak, R.; Mann, K.; Kuehn, K.  
EMBO J. 12, 4795-4802, 1993  
A;Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen n  
A;Reference number: S39614; MUID:94038963; PMID:8223488  
A;Accession: S39614  
A;Molecule type: Protein  
A;Residues: 371-554 <EBL>  
R;Babel, W.; Glanville, R.W.  
Eur. J. Biochem. 143, 545-556, 1984  
A;Title: Structure of human-basement-membrane (type IV) collagen. Complete amino-acid seq  
A;Reference number: A02863; MUID:85003629; PMID:6434307



A:Accession: A02863  
A:Molecule type: protein  
A:Residues: 534-718, D', 720-836, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'K', 999-  
A:Experimental source: Placenta  
R:Glanville, R.W.; Rauter, A.  
Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981  
A:Title: Pepsin fragments of human placental basement-membrane collagens showing interrupted  
A:Reference number: S16908; MUID:82005835; PMID:6792033  
A:Accession: A58517  
A:Molecule type: protein  
A:Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-553, 1389-1405, 'XX', 1408-1409, 'X', 1411-14  
R:Macwright, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.  
Biochemistry 22, 4940-4948, 1983  
A:Title: Isolation and characterization of pepsin-solubilized human basement membrane (b  
A:Reference number: S16910; MUID:84053346; PMID:6416291  
A:Accession: S16910  
A:Molecule type: protein  
A:Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-549, 939-940, 'M', 942-944, 'Y', 946, 'X', 948-  
A:Experimental source: Placenta  
R:Pihlajaniemi, T.; Tryggvason, K.; Myers, J.C.; Kurkinen, M.; Lebo, R.; Cheung, M.C.; F  
J. Biol. Chem. 260, 7681-7687, 1985  
A:Title: cDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen  
A:Reference number: S01466; MUID:85207819; PMID:2581969  
A:Accession: S01466  
A:Molecule type: mRNA  
A:Residues: 1256-1669 <PIH>  
A:Cross-references: EMBL:M10940; NID:G180421; PIDN:AAA52006.1; PID:G180424  
R:Brinker, J.M.; Gudas, L.J.; Loidl, H.R.; Wang, S.Y.; Rosenbloom, J.; Kefalides, N.A.;  
Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985  
A:Title: Restricted homology between human alpha-1 type IV and other procollagen chains.  
A:Reference number: S16879; MUID:85216555; PMID:2582422  
A:Accession: S16879  
A:Molecule type: mRNA  
A:Residues: 1259-1669 <BRI>  
A:Cross-references: EMBL:M11315; NID:G180817; PIDN:AAA52042.1; PID:G180818  
R:Oberbauer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss,  
Eur. J. Biochem. 147, 217-224, 1985  
A:Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1  
A:Reference number: A02864; MUID:85127033; PMID:2578961  
A:Accession: S19091  
A:Molecule type: protein  
A:Residues: 1435-1461, 'H', 1463-1482, 'X', 1484-1491, 1501-1514, 'X', 1516-1519, 1534-1553, 'X',  
R:Siebold, B.; Deutzmann, R.; Kuehn, K.  
Eur. J. Biochem. 176, 617-624, 1988  
A:Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxyterm  
A:Reference number: S02550; MUID:89005112; PMID:2844531  
A:Contents: annotation; disulfide bonds  
C:Genetics:  
A:Gene: GDB:COL4A1  
A:Cross-references: GDB:119791; OMIM:120130  
A:Map position: 13q34-13q34  
A:Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3; 147/3; 156/3; 184/3; 205/3; 217/3; 231/  
/1; 731/3; 782/1; 820/1; 876/1; 906/1; 957/1; 990/1; 1020/1; 1066/3; 1109/1; 1136/1; 118  
C:Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha 2(  
ociations among trimer amino-terminal domains (disulfide and desmosine cross-links), dim  
r-trimer associations in the interrupted helical domain (with disulfide and desmosine cr  
C:Function:  
A:Description: structural component of extracellular basement membrane  
A:Superfamily: collagen alpha 1(IV) chain  
C:Keywords: basement membrane; blocked amino end; cell binding; coiled coil; duplication  
F:1-26/Domain: signal sequence #status predicted <SIG>  
F:27-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>  
F:29-162/Domain: amino-terminal nonhelical, 7S <7SD>  
F:163-1440/Domain: interrupted helical <COL>  
F:414-452/Region: integrin binding #status experimental  
F:597-599/Region: cell attachment (R-G-D) motif  
F:917-919/Region: cell attachment (R-G-D) motif  
F:968-970/Region: cell attachment (R-G-D) motif  
F:1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>  
F:1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>  
F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>  
F:27/Modified site: blocked amino end (A1a) (in mature form) #status experimental  
A:Reference number: S125, 434, 467, 470/disulfide bonds: interchain #status predicted

F:45, 48, 78, 90, 129, 156, 172, 217, 228, 231, 277, 295, 298, 322, 343, 361, 460, 463, 497, 527, 540, 543, 573,  
1081, 1084, 1099, 1117, 1132, 1150, 1165, 1182, 1185, 1188, 1206, 1235, 1255, 1283, 1304, 1319, 1328, 1341,  
F:45, 48, 78, 90, 129, 156, 217, 228, 231, 277, 295, 298, 322, 343, 361, 460, 463, 497, 527, 543, 573, 582, 61,  
99, 1117, 1132, 1150, 1165, 1182, 1185, 1188, 1206, 1235, 1265, 1283, 1304, 1319, 1328, 1340, 1356, 1371, 3  
F:54, 63, 75, 84, 87, 96, 102, 105, 108, 111, 117, 120, 123, 138, 141, 147, 150, 153, 159, 167, 178, 181, 184,  
419, 422, 425, 439, 445, 448, 451, 479, 485, 491, 494, 503, 512, 518, 524, 530, 546, 549, 552, 555, 561, 567,  
9, 745, 748, 751, 754, 763/Modified site: 4-hydroxyproline (Pro) #status experimental  
F:126/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F:129/Modified site: allysine (Lys) #status predicted  
F:172, 540, 547/Modified site: 5-hydroxylysine (Lys) #status atypical  
F:272, 645, 839/Modified site: 4-hydroxyproline (Pro) #status atypical  
F:446-447/Cleavage site: Gly-Ile (gelatinase B) #status predicted  
F:766, 775, 784, 787, 790, 796, 799, 804, 810, 816, 822, 834, 860, 863, 869, 872, 875, 887, 890, 893, 899, 902,  
23, 1129, 1138, 1141, 1159, 1171, 1176, 1179, 1194, 1200, 1203, 1215, 1224, 1227, 1244, 1247, 1250, 1256, 1  
431, 1437/Modified site: 4-hydroxyproline (Pro) #status experimental  
F:1120, 1268/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental  
F:1120, 1268/Binding site: carbohydrate (Lys) (covalent) (partial) #status experimental  
F:1214, 1424/Modified site: 3-hydroxyproline (Pro) #status absent  
F:1392, 1395, 1398, 1404/Modified site: 4-hydroxyproline (Pro) #status experimental  
F:1460-1548, 1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted  
F:1505-1511, 1616-1622/Disulfide bonds: #status predicted  
F:1570-1662, 1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Query Match 15.4%; Score 70; DB 1; Length 1669;  
Best Local Similarity 31.1%; Pred. No. 31;  
Matches 23; Conservative 7; Mismatches 24; Indels 20; Gaps 3;  
QY 10 AGCGFADAHWT-----GLWTGLG---EGEGGGGPGEGQASPTPD-----CASR 49  
Db 34 SCGCKCDDHGKGGKGERGLGQGVGPGMGEGPGPGQKQKDTGEPGLPTGKGR 93  
QY 50 WPRSASRWPMWSAGL 63  
Db 94 GPPGASGYPGNPGL 107

RESULT 5  
T35294  
Probable endo alpha-1,4 polygalactosaminidase - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
R:Seeger, K.; Harris, D.; Bentley, S.D.; Barkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, July 1999  
A:Reference number: Z21574  
A:Accession: T35294  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-282 <SEE>  
A:Cross-references: EMBL:AL096872; PIDN:CAB51262.1; GSPDB:GN00070; SCOEDB:SC5F7.23c  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SC5F7.23c

Query Match 14.9%; Score 68; DB 2; Length 282;  
Best Local Similarity 37.3%; Pred. No. 9;  
Matches 19; Conservative 3; Mismatches 25; Indels 4; Gaps 2;  
QY 10 AGCGFADAHWTGLWTGLGEGGGGPGEGQASPTPDCA SRW-PRASRWPMW 59  
Db 16 AGCTAPGDDGGVGGVGGEGGGGGGAT---AGHWRPTGTANQW 63

RESULT 6  
B45344  
Probable nuclear antigen - suid herpesvirus 1 (strain Kaplan)  
C:Species: suid herpesvirus 1  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Jul-1999  
R:Vlcek, C.; Kozmik, Z.; Paces, V.; Schirm, S.; Schwyzler, M.  
Virology 179, 365-377, 1990  
A:Title: Pseudorabies virus immediate-early gene overlaps with an oppositely oriented op  
A:Reference number: A45344; MUID:91021039; PMID:2171211

```

Best Local Similarity   29.8%; Pred. No. 17;
Matches    14;      Conservative    6; Mismatches    25; Indels    2; Gaps    1;

QY    31 EGGIGPEGOASTPTPCASRWPP--RSASRWPWSAGLTVDRPQLGELC 75
      :|||:::|||||:|||||:|||||:|||||:|||||:|||||:
DB    343 KGGPGRGCGCLPDATHNFFVRQLVTNTGWGTGLDIRFNGIGHPC 389

RESULT 9
A$4411
amine oxidase (copper-containing) (EC 1.4.3.6), serum, precursor - bovine
C$Species: Bos primigenius taurus (cattle)
C$Date: 09-Sep-1994 #sequence revision 06-Feb-1995 #text_change 18-Feb-2000
C$Accession: A$4411; B$number: A48242; S$65408
R$Nu, D$: Medzhitarsky, K.F.; Adams, G.W.; Mayer, P.; Hines, W.M.; Burlingame, A.L.; Smi
J. Biol. Chem. 269, 9326-9332, 1994
A$Title: Primary structures for a mammalian cellular and serum copper amine oxidase.
A$Reference number: A$4411; MUID:94193686; PMID:8144587
A$Accession: A$4411
A:Molecule type: mRNA
A$Residues: 1-762 <MUA>
A$Cross-references: GB:S65983; NID:9546215; PIDN:AAB30397.1; PID:9546216
R$Nu, D.: Janes, S.M.; Smith, A.J.; Brown, D.E.; Dooley, D.M.; Klinman, J.P.
J. Biol. Chem. 267, 7979-7982, 1992
A$Title: Tyrosine codon corresponds to topa quinone at the active site of copper amine oxi
A$Reference number: A$38081; MUID:92235001; PMID:1569055
A$Accession: B$38081
A:Molecule type: protein
A$Residues: 463-465, 'D', 467-473, 'X', 475-485 <MU2>
R$Nu, D.: Mu, D.; Wenner, D.; Smith, A.J.; Kaur, S.; Maltby, D.; Burlingame, A.L.; F
Science 248, 981-987, 1990
A$Title: A new redox cofactor in eukaryotic enzymes: 6-hydroxydopa at the active site of
A$Reference number: A$8242; MUID:90260648; PMID:2111581
A$Accession: A$8242
A>Status: preliminary
A:Molecule type: protein
A$Residues: 463-469, 'X', 471-487 <DEB>
C$Superfamily: amiloride-binding protein
C$Keywords: copper; glycoprotein; oxidoreductase; quinoxaline; topaquinine
F$1-16/Domain: signal sequence #status predicted <SIG>
F$17-762/Product: amine oxidase (copper-containing) #status predicted <MAT>
F$136,231,685/Binding site: carbohydrate (Asn) (covalent) #status predicted
F$470/Modified site: topaquione (Iyr) #status experimental

Query Match          14.7%; Score 67; DB 2; Length 762;
Best Local Similarity 38.6%; Pred. No. 30;
Matches    22;      Conservative    3; Mismatches    18; Indels    14; Gaps    3;

QY    22 LWTGL-----GEQEGGIGPEGOASTPTPCASRWPSASRWPWSAGLTVDRPQL 71
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB    9 LMTLLVMGREGVSGEYGKQHCHPSLPPCRSP---SDQPW----THPDSQL 58

RESULT 10
T00027
brain-specific angiogenesis inhibitor 2 - human
N$Alternate names: BA12 protein
C$Species: Homo sapiens (man)
C$Date: 22-Jan-1999 #sequence revision 22-Jan-1999 #text_change 21-Jul-2000
C$Accession: T00027
R$Shitarauchi, T.; Nishimori, H.; Ichise, H.; Nakamura, Y.; Tokino, T.
Cytochrome Cell Genet. 79, 103-108, 1997
A$Title: Cloning and characterization of BA12 and BA13 novel genes homologous to brain-
A$Reference number: Z14066; MUID:98194217; PMID:9533023
A$Accession: T00027
A>Status: translated from GB/EMBL/DBJ

```

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-740, 'B', 712-1154 <CA2>  
A:Cross-references: EMBL:L25663; NID:G409762; PIDN:AAA21792.1; PID:G409764  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1993  
C:Genetics:  
A:Mobile element: retrotransposon Tad1-1

Query Match 14.5%; Score 66; DB 2; Length 1154;  
Best Local Similarity 28.4%; Pred. No. 56;  
Matches 19; Conservative 1; Mismatches 37; Indels 10; Gaps 2;

QY 3 GRPRVRSAGCGFADAHWTGLWT-----GLGEGGEGGIGPEGQASPTPDCASTR--WPR 52  
|||  
DB 910 GRPPADKPSSGAKAKAEAWAPTEKLLANAGAPPPESCTPPASQIPEPTAADREWNLE 969  
|||

QY 53 SASRNPW 59  
|:  
DB 970 DATNREW 976  
|:

RESULT 13  
AG1940  
hypothetical protein alr1074 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Residues: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C:Accession: AG1940  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AG1940  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-219 <XR>  
A:Cross-references: GB:BA000019; PIDN:BAF73031.1; PID:G17130420; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr1074  
C:Superfamily: Synchocystis hypothetical protein sir0885

Query Match 14.4%; Score 65.5; DB 2; Length 219;  
Best Local Similarity 31.7%; Pred. No. 13;  
Matches 19; Conservative 7; Mismatches 23; Indels 11; Gaps 4;

QY 21 GLW-----TGLGEGGEGGIGPEGQASPTPDCASTRWPSASRWPSAGLTVRDRPQLGELC 75  
|||  
DB 89 GLWTKQDTGVGGSSGGFTLNGAVRSPDAA--WIKKA-RW---EAIPEQRKRFAPIC 142  
|||

RESULT 14  
T30630  
hypothetical protein 28L - Molluscum contagiosum virus 1  
N:Alternate names: MC028L  
C:Species: Molluscum contagiosum virus 1  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 11-May-2000  
C:Accession: T30630  
R:Senkevich, T.G.; Bugert, J.J.; Sisler, J.R.; Koonin, E.V.; Darai, G.; Moss, B.  
Science 273, 813-816, 1996  
A:Title: Genome sequence of a human tumorigenic poxvirus: Prediction of specific host re  
A:Reference number: Z20876; MUID:96325459; PMID:8670425  
A:Accession: T30630  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-268 <SEN>  
A:Cross-references: EMBL:U60315; PIDN:AAC55156.1  
C:Genetics:  
A>Note: MC028L

Query Match 14.4%; Score 65.5; DB 2; Length 268;  
Best Local Similarity 38.3%; Pred. No. 16;

Mon Aug 16 17:45:51 2004

Matches 18; Conservative 6; Mismatches 18; Indels 5; Gaps 2;

QY 25 GLGEGQEGGI---GPEGQASPTDCASR-WPRSASRWPSAGLTVR 66  
 DB 27 GAGSASEDDAEPGPEFPGVPEPAARSPRAARLPAGAGLPRR 73

RESULT 15

A24925  
 beta-galactosidase (EC 3.2.1.23) - Klebsiella pneumoniae  
 C:Species: Klebsiella pneumoniae  
 C:Date: 23-Aug-1987 #sequence\_revision 23-Aug-1987 #text\_change 22-Jun-1999  
 C:Accession: A24925  
 R:Buvinger, W.E.; Riley, M.  
 J. Bacteriol. 163, 850-857, 1985  
 A:Title: Nucleotide sequence of Klebsiella pneumoniae lac genes.  
 A:Reference number: A91803; MUID:85289025; PMID:3897196  
 A:Accession: A24925  
 A:Molecule type: DNA  
 A:Residues: 1-1034 <BUV>  
 A:Cross-references: GB:M11441; GB:M11416; NID:G149216; PIDN:AAA25082.1; PID:G149218  
 C:Genetics:  
 A:Gene: lacZ  
 C:Superfamily: beta-galactosidase  
 C:Keywords: glycosidase; hydrolase

Query Match 14.4%; Score 65.5; DB 2; Length 1034;  
 Best Local Similarity 32.7%; Pred. No. 57;  
 Matches 17; Conservative 6; Mismatches 20; Indels 9; Gaps 3;

QY 18 HW-----TGLWTGLGEGQEGGIGPEGQASPTDCASRWPSASRWPSAGL 63  
 DB 983 HWHKMQAEDGVWITL-DGLHGVG--GDDSWTPSVLPWLLSQTRWQYEVSL 1031

Search completed: August 13, 2004, 09:07:51  
 Job time : 18 secs

GenCore version 5.1.6

Copyright (C) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2004, 08:56:32 ; Search time 13 Seconds

(without alignments)

316.426 Million cell updates/sec

Title: US-09-972-032-2

Perfect score: 456

Sequence: 1 MCGRRRVRSAGCGFADAHWT.....SAGLTVRDRPQIGELCMGRG 79

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	81	17.8	367	1 LHX8 MOUSE	Q35652 mus musculus
2	73.5	16.1	1729	1 TABP_HUMAN	Q9C0C2 homo sapien
3	70	15.4	1669	1 CA14_HUMAN	P02462 homo sapien
4	68.5	15.0	310	1 NF1U_AZOBK	Q43909 azospirillum
5	67.5	14.8	1733	1 VNDA_PVRKA	P23485 pseudorabies
6	67	14.7	333	1 NK32_HUMAN	P78367 homo sapien
7	67	14.7	547	1 RM56_HUMAN	P83111 homo sapien
8	67	14.7	762	1 AOCX_BOVIN	Q29437 bos taurus
9	67	14.7	762	1 AOCY_BOVIN	O46406 bos taurus
10	67	14.7	1572	1 BA12_HUMAN	O62241 homo sapien
11	66	14.5	1003	1 TF3A_MOUSE	O70157 mus musculus
12	65.5	14.4	1034	1 BGAL_KLEPN	P06219 klebsiella
13	64.5	14.1	1669	1 CA14_MOUSE	P02463 mus musculus
14	64	14.0	576	1 Z384_HUMAN	Q8tf68 homo sapien
15	63.5	13.9	275	1 TRVA_HUMAN	P15157 homo sapien
16	63.5	13.9	596	1 FIBA_BOVIN	P02672 bos taurus
17	63	13.8	376	1 FRAE_BOEPE	Q00879 bordetella
18	63	13.8	579	1 Z384_RAT	Q09eq4 rattus norv
19	63	13.8	1216	1 AEGP_RAT	Q63191 rattus norv
20	63	13.8	1838	1 CHA2_HUMAN	P20908 homo sapien
21	62	13.6	132	1 CHA2_BOMMO	P08825 bombyx mori
22	62	13.6	426	1 EGL1_HUMAN	Q9gz19 homo sapien
23	62	13.6	485	1 ONC2_HUMAN	P95948 homo sapien
24	61.5	13.5	745	1 ATCS_SYNY3	P73241 synechocyst
25	61	13.4	651	1 CSP6_HUMAN	Q9nv66 h cofactor
26	61	13.4	1139	1 M2A2_HUMAN	P49641 homo sapien
27	61	13.4	1171	1 GLG1_RAT	Q62638 rattus norv
28	60.5	13.3	164	1 LWA_ACTEO	Q16998 actinia equ
29	60.5	13.3	389	1 WN1E_MOUSE	P48614 mus musculus
30	60.5	13.3	520	1 CMCH_NOCLA	Q51080 noxia la
31	60	13.2	161	1 CHB2_BOMMO	P08828 bombyx mori
32	60	13.2	226	1 NUKM_NEUCR	O47950 neurospora
33	60	13.2	616	1 RFX5_HUMAN	P48382 homo sapien

## ALIGNMENTS

## RESULT 1

ID	LHX8_MOUSE	STANDARD;	PRT;	367 AA.
AC	O35652; 070163; 088707;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	LIM/homeobox protein Lhx8 (LIM homeodomain Lhx7) (L3).			
GN	LHX8 OR LHX7			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Chordata; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ICR; TISSUE=Brain;			
RX	MEDLINE=97083417; PubMed=8929991;			
RA	Matsumoto K., Tanaka T., Furuyama T., Kashiwara Y., Mori T., Ishii N.,			
RA	Kitanaka J., Takemura M., Tohyama M., Wanaoka A.;			
RT	"I3, a novel murine LIM-homeodomain transcription factor expressed in			
RT	the ventral telencephalon and the mesenchyme surrounding the oral			
RT	cavity.";			
RL	Neurosci. Lett. 204:113-116(1996).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129/SVJ; TISSUE=Liver;			
RX	MEDLINE=98260681; PubMed=9598319;			
RA	Kitanaka J.-I., Takemura M., Matsumoto K., Mori T., Wanaoka A.;			
RT	"Structure and chromosomal localization of a murine LIM/homeobox gene,			
RT	Lhx8.";			
RL	Genomics 49:307-309(1998).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	MEDLINE=98237706; PubMed=9570771;			
RA	Grigoriou M., Tucker A.S., Sharpe P.T., Pachnis V.;			
RT	"Expression and regulation of Lhx6 and Lhx7, a novel subfamily of LIM			
RT	homeodomain encoding genes, suggests a role in mammalian head			
RT	development.";			
RL	Development 125:2063-2074(1998).			
CC	-I- FUNCTION: Transcription factor involved in differentiation of			
CC	certain neurons and mesenchymal cells.			
CC	-I- SUBCELLULAR LOCATION: Nuclear (Probable).			
CC	-I- DEVELOPMENTAL STAGE: Limited spatially to the medical ganglionic			
CC	emience and the mesenchyme surrounding the oral cavity and			
CC	temporarily from middle embryonic to early postnatal development.			
CC	-I- SIMILARITY: Contains 1 homeobox domain.			
CC	-I- SIMILARITY: Contains 2 LIM zinc-binding domains.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation-			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			

P39881 canis famill  
 Q95076 homo sapien  
 Q9bzz6 homo sapien  
 P06475 herpes simp  
 Q89730 herpes simp  
 P08933 sus scrofa  
 Q9xsn2 ovis aries  
 P08127 pseudomonas  
 P06622 pseudomonas  
 P27887 pseudomonas  
 P23490 homo sapien  
 P38059 pseudomonas

34 60 13.2 975 1 CUTI\_CANFA  
 35 59.5 13.0 343 1 ALX3\_HUMAN  
 36 59.5 13.0 473 1 RT4R\_HUMAN  
 37 59.5 13.0 480 1 VGLC\_HSV23  
 38 59.5 13.0 480 1 VGLC\_HSV2H  
 39 59.5 13.0 926 1 PERP\_PIG  
 40 59 12.9 273 1 TRYI\_SHEEP  
 41 59 12.9 307 1 NAHH\_PSEPU  
 42 59 12.9 307 1 XYLE\_PSEPU  
 43 59 12.9 307 1 XYLE\_PSEAE  
 44 59 12.9 316 1 LORI\_HUMAN  
 45 59 12.9 384 1 CARA\_PSEST



```
FT CONFLICT 1450 1450 F -> S (IN REF. 2).
SQ SEQUENCE 1729 AA; 181814 MW; C65F38FA37045C4A CRC64;

Query Match
Best Local Similarity 27.6%; Score 73.5; DB 1; Length 1729;
Matches 24; Conservative 4; Mismatches 30; Indels 29; Gaps 2;

QY 6 RVVSGCGFADAHWTGLTGLTGEGGEGGIG-----PEGQASPTPD 45
DB 1470 RESAASGLG-----GLLEEGCAGAGAAQEEVLEPGRDSPSPWRPQDGEASQTED 1520
QY 46 CASRWPRGASRWPSAGLTVDRPQLG 72
DB 1521 VDGWTGSSAARWSDGPAQTSRPSQG 1547

RESULT 3
CA14 HUMAN
ID CA14 HUMAN STANDARD; PRT; 1669 AA.
AC P02462;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(IV) chain precursor.
GN COL4A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89340433; PubMed=2701944;
RA Soaninen R., Huotari M., Ganguly A., Prockop D.J., Tryggvason K.;
RT "Structural organization of the gene for the alpha 1 chain of human
RT type IV collagen."
RL J. Biol. Chem. 264:13565-13571(1989).
RN [2]
RP SEQUENCE OF 46-1257 FROM N.A.
RX TISSUE=Placenta;
RX MEDLINE=88083584; PubMed=3691802;
RA Soaninen R., Haka-Risku T., Prockop D.J., Tryggvason K.;
RT "Complete primary structure of the alpha 1-chain of human basement
RT membrane (type IV) collagen."
RL FEBS Lett. 225:188-194(1987).
RN [3]
RP SEQUENCE OF 1-943 FROM N.A.
RX TISSUE=Placenta;
RX MEDLINE=86029471; PubMed=3311751;
RA Brazel D., Oberbauer I., Dieringer H., Babel W., Glanville R.W.,
RA Deutzmann R., Kuehn K.;
RT "Completion of the amino acid sequence of the alpha 1 chain of human
RT basement membrane collagen (type IV) reveals 21 non-triplet
RT interruptions located within the collagenous domain."
RL Eur. J. Biochem. 168:523-536(1987).
RN [4]
RP SEQUENCE OF 28-243.
RX MEDLINE=86004708; PubMed=4043082;
RA Glanville R.W., Qian R.Q., Siebold B., Risteli J., Kuehn K.;
RT "Amino acid sequence of the N-terminal aggregation and cross-linking
RT region (7S domain) of the alpha 1 (IV) chain of human basement
RT membrane collagen."
RL Eur. J. Biochem. 152:213-219(1985).
RN [5]
RP SEQUENCE OF 534-1447.
RX MEDLINE=85003629; PubMed=6434307;
RA Babel W., Glanville R.W.;
RT "Structure of human basement-membrane (type IV) collagen. Complete
RT amino-acid sequence of a 914-residue-long pepsin fragment from the
RT alpha 1(IV) chain."
RL Eur. J. Biochem. 143:545-556(1984).
RN [6]
RP SEQUENCE OF 1256-1669 FROM N.A.
RX MEDLINE=85207819; PubMed=2581969;
```

```
RA Pihlajaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,
RA Cheung M.-C., Prockop D.J., Boyd C.D.;
RT "cDNA clones coding for the pro-alpha1(IV) chain of human type IV
RT procollagen reveal an unusual homology of amino acid sequences in two
RT halves of the carboxyl-terminal domain."
RL J. Biol. Chem. 260:7681-7687(1985).
RN [7]
RP SEQUENCE OF 1259-1669 FROM N.A.
RX MEDLINE=85216555; PubMed=2582422;
RA Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,
RA Kefalides N.A., Myers J.C.;
RT "Restricted homology between human alpha 1 type IV and other
RT procollagen chains."
RL Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).
RN [8]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89034231; PubMed=3182844;
RA Soaninen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;
RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
RT collagen are divergently encoded on opposite DNA strands and have an
RT overlapping promoter region."
RL J. Biol. Chem. 263:17217-17220(1988).
RN [9]
RP SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.
RX TISSUE=Placenta;
RX MEDLINE=89005112; PubMed=2844531;
RA Siebold B., Deutzmann R., Kuehn K.;
RT "The arrangement of intra- and intermolecular disulfide bonds in the
RT carboxyterminal, non-collagenous aggregation and cross-linking domain
RT of basement-membrane type IV collagen."
RL Eur. J. Biochem. 176:617-624(1988).
RN [10]
RP FUNCTION: Type IV collagen is the major structural component of
RN glomerular basement membranes (GBM), forming a 'chicken-wire'
RN meshwork together with laminins, proteoglycans and entactin/
RN nidogen.
RN [11]
RP SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
RN alpha 6(IV), each of which can form a triple helix structure
RN with 2 other chains to generate type IV collagen network.
RN [12]
RP DOMAIN: Alpha chains of type IV collagen have a noncollagenous
RN domain (NC1) at their C-terminus, frequent interruptions of the
RN G-X-Y repeats in the long central triple-helical domain (which may
RN cause flexibility in the triple helix), and a short N-terminal
RN triple-helical 7S domain.
RN [13]
RP PTM: Lysines at the third position of the tripeptide repeating
RN unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.
RN [14]
RP PTM: Prolines at the third position of the tripeptide repeating
RN unit (G-X-Y) are hydroxylated in some or all of the chains.
RN [15]
RP PTM: Type IV collagens contain numerous cysteine residues which
RN are involved in inter- and intramolecular disulfide bonding. 12 of
RN these, located in the NC1 domain, are conserved in all known type
RN IV collagens.
RN [16]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See http://www.isb-sib.ch/announce/
RN or send an email to license@isb-sib.ch).
RN [17]
RP EMBL; M26576; AAA53098.1; JOINED.
RN EMBL; J04217; AAA53098.1; JOINED.
RN EMBL; M26550; AAA53098.1; JOINED.
RN EMBL; M26540; AAA53098.1; JOINED.
RN EMBL; M26542; AAA53098.1; JOINED.
RN EMBL; M26543; AAA53098.1; JOINED.
RN EMBL; M26544; AAA53098.1; JOINED.
RN EMBL; M26545; AAA53098.1; JOINED.
RN EMBL; M26546; AAA53098.1; JOINED.
RN EMBL; M26547; AAA53098.1; JOINED.
RN EMBL; M26537; AAA53098.1; JOINED.
RN EMBL; M26538; AAA53098.1; JOINED.
RN EMBL; M26548; AAA53098.1; JOINED.
```

```
DR EMBL; M26549; AAA53098.1; JOINED.
DR EMBL; M26551; AAA53098.1; JOINED.
DR EMBL; M26552; AAA53098.1; JOINED.
DR EMBL; M26553; AAA53098.1; JOINED.
DR EMBL; M26554; AAA53098.1; JOINED.
DR EMBL; M26555; AAA53098.1; JOINED.
DR EMBL; M26556; AAA53098.1; JOINED.
DR EMBL; M26557; AAA53098.1; JOINED.
DR EMBL; M26558; AAA53098.1; JOINED.
DR EMBL; M26559; AAA53098.1; JOINED.
DR EMBL; M26560; AAA53098.1; JOINED.
DR EMBL; M26561; AAA53098.1; JOINED.
DR EMBL; M26562; AAA53098.1; JOINED.
DR EMBL; M26563; AAA53098.1; JOINED.
DR EMBL; M26564; AAA53098.1; JOINED.
DR EMBL; M26565; AAA53098.1; JOINED.
DR EMBL; M26566; AAA53098.1; JOINED.
DR EMBL; M26567; AAA53098.1; JOINED.
DR EMBL; M26568; AAA53098.1; JOINED.
DR EMBL; M26569; AAA53098.1; JOINED.
DR EMBL; M26570; AAA53098.1; JOINED.
DR EMBL; M26571; AAA53098.1; JOINED.
DR EMBL; M26572; AAA53098.1; JOINED.
DR EMBL; M26573; AAA53098.1; JOINED.
DR EMBL; M26574; AAA53098.1; JOINED.
DR EMBL; M26575; AAA53098.1; JOINED.
DR EMBL; M26576; AAA53098.1; JOINED.
DR EMBL; X05561; CAA29075.1; -.
DR EMBL; M10940; AAA52006.1; -.
DR EMBL; M11315; AAA52042.1; -.
DR PIR; S16876; CGHU4B.
DR Genew; HGNC:2202; COL4A1.
DR MIM; 120130; -.
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 24.
DR ProDom; PD000007; C1g helix; 6.
DR ProDom; PD003923; ProcollagenC4; 1.
DR SMART; SM00111; C4; 2.
DR Extracellular matrix; Connective tissue; Basement membrane;
KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT SIGNAL 1 27
FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .).
FT DISULFID 1460 1551 OR 1548.
FT DISULFID 1493 1548 OR 1551.
FT DISULFID 1505 1511 OR 1662.
FT DISULFID 1570 1665 OR 1665.
FT DISULFID 1604 1662 OR 1665.
FT DISULFID 1616 1622
FT CONFLICT 227 238 SG -> KE (IN REF. 4).
FT CONFLICT 241 241 G -> K (IN REF. 4).
FT CONFLICT 319 319 N -> D (IN REF. 3).
FT CONFLICT 719 719 N -> D (IN REF. 5).
FT CONFLICT 837 837 D -> Y (IN REF. 5).
FT CONFLICT 842 842 K -> P (IN REF. 5).
FT CONFLICT 896 896 V -> W (IN REF. 2).
FT CONFLICT 896 896 E -> Q (IN REF. 5).
FT CONFLICT 904 904 S -> K (IN REF. 5).
FT CONFLICT 944 944 S -> K (IN REF. 5).
FT CONFLICT 998 998 K -> P (IN REF. 5).
FT CONFLICT 1010 1010 K -> P (IN REF. 5).
FT CONFLICT 1012 1012 E -> Q (IN REF. 5).
FT CONFLICT 1358 1358 E -> Q (IN REF. 5).
FT CONFLICT 1669 AA; 160611 MW; 3BBA6DFFB9B8A84 CRC64;
SQ SEQUENCE 1669 AA; 160611 MW; 3BBA6DFFB9B8A84 CRC64;

Query Match 15.4%; Score 70; DB 1; Length 1669;
Best Local Similarity 31.1%; Pred. NO. 25;
Matches 23; Conservative 7; Mismatches 24; Indels 20; Gaps 3;

QY 10 AGCGFADAHWT-----GLWTGLG-----EGOEGGIGPEGQASPTPD-----CASR 49
Db 34 SCGKCDCHGVGQKGERGLPGLQGVIFPGMQGPEGPGPGKGDIGELPCTKCTR 93
QY 50 WPSASRWPWSAGL 63
Db 94 GPPGASGYGNGDL 107

RESULT 4
NIFU AZOBR
ID NIFU AZOBR STANDARD; PRT; 310 AA.
AC Q4309; P70726;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Nitrogen fixation protein nifu.
GN NIFU.
OS Azospirillum brasilense.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Rhodospirillaceae; Azospirillum.
OC NCBI_TaxID=192;
RN [1]
SEQUENCE FROM N.A.
RA Frazzon J.S., Schrank I.S.;
RT "The nifu gene from Azospirillum brasilense.";
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE=98164354; PubMed=9503607;
RA Frazzon J.S., Schrank I.S.;
RT "Sequencing and complementation analysis of the nifusv genes from
AZospirillum brasilense.";
RL FEMS Microbiol. Lett. 159:151-157(1998).
CC -!- FUNCTION: MAY BE INVOLVED IN THE FORMATION OR REPAIR OF [FE-S]
CC -!- CLUSTERS PRESENT IN IRON-SULFUR PROTEINS (POTENTIAL).
CC -!- SIMILARITY: Belongs to the nifu family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; L34354; AAA22184.1; -.
CC EMBL; U26437; AAC46176.1; -.
CC InterPro; IPR007419; Fer2_BFD.
CC InterPro; IPR001075; NifU_C.
CC InterPro; IPR002871; NifU_N.
CC Pfam; PF04324; fer2_BFD; 1.
CC Pfam; PF01106; NifU-like; 1.
CC Pfam; PF01592; NifU_N; 1.
CC ProDom; PD002830; NifU_C; 1.
CC Nitrogen fixation.
KW CONFLICT 56
FT CONFLICT 133 136 N -> K (IN REF. 1).
FT CONFLICT 141 144 ESSE -> AAGT (IN REF. 1).
FT CONFLICT 149 152 CFGI -> DRQA (IN REF. 1).
FT CONFLICT 201 202 IERA -> DRRR (IN REF. 1).
FT CONFLICT 212 223 MISSING (IN REF. 1).
FT CONFLICT 230 231 TPSSRRPRRH -> IAYQAAGPKAAE (IN REF. 1).
FT CONFLICT 235 236 NV -> TC (IN REF. 1).
FT CONFLICT 243 244 OK -> RR (IN REF. 1).
FT CONFLICT 281 282 DW -> EL (IN REF. 1).
FT CONFLICT 303 307 GT -> AS (IN REF. 1).
FT CONFLICT 310 AA; 33201 MW; EF9F474F91255FEF CRC64;
SQ SEQUENCE 310 AA; 33201 MW; EF9F474F91255FEF CRC64;
```



Query Match 15.0%; Score 68.5; DB 1; Length 310;  
Best Local Similarity 51.9%; Pred.No.7.3;  
Matches 14; Conservative 1; Mismatches 11; Indels 1; Gaps 1;

QY 32 GGIGECQAQPPTDCASRWPRASRWP 58  
DB 193 GAVGPAQAPSPTPPARSGWIPS-SRWP 218

RESULT 5  
VNUA PRVKA STANDARD; PRT; 1733 AA.  
ID VNUA PRVKA  
AC P33485;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-FEB-1994 (Rel. 28, Last annotation update)  
DE Probable nuclear antigen.  
OS Pseudorabies virus (strain Kaplan) (PRV).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Varicellovirus.  
NCBI\_TaxID=33703;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91021039; PubMed=2171211;  
RA Vacek C., Kozmik Z., Faces V., Schilm S., Schwyzler M.;  
RT "Pseudorabies virus immediate-early gene overlaps with an oppositely  
RT oriented open reading frame: characterization of their promoter and  
RT enhancer regions".  
RL Virology 179:365-377(1990).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).

EMBL; M34651; AAA47471.1; --  
PIR; B45344; B45344.  
FT DOMAIN 112 117 POLY-THR.  
FT DOMAIN 179 1733 GLY-RICH.  
FT DOMAIN 192 196 POLY-SER.  
FT DOMAIN 271 298 POLY-PRO.  
FT DOMAIN 304 308 POLY-ARG.  
FT DOMAIN 883 889 POLY-GLY.  
FT DOMAIN 1398 1405 POLY-GLY.

SQ SEQUENCE 1733 AA; 172166 MW; 0C8CD8BE475BB5E2 CRC64;

Query Match 14.8%; Score 67.5; DB 1; Length 1733;  
Best Local Similarity 33.8%; Pred.No.46;  
Matches 25; Conservative 3; Mismatches 39; Indels 7; Gaps 2;

QY 8 VSAGCGFADAHWTGLWGEGGGTGPGQASPTEPCASRWP--RSASRWPSAGLTV 65  
DB 967 VAGGAGEA----GLNGAGLGGAGALCAGGAGCGGAGCGGARRRRRWDEAGLLG 1021

QY 66 RDRPQLGELCMGRG 79  
DB 1022 PERQGAGRGLRGPG 1035

RESULT 6  
NK32 HUMAN STANDARD; PRT; 333 AA.  
ID NK32 HUMAN  
AC P78367;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Homeobox protein NKx-3.2 (Bagpipe homeobox protein homolog 1).  
GN BAPX1 OR NKX3B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



DR	PRINTS; PR00766; CUDAOXIDASE.
DR	PROSITE; PS01164; COPPER AMINE OXID 1; 1.
DR	PROSITE; PS01165; COPPER AMINE OXID 2; 1.
KW	Oxidoreductase; Copper; TPQ; Glycoprotein; Signal.
FT	SIGNAL 1 16
FT	CHAIN 17 762
FT	MOD_RES 470 470
FT	METAL 519 519
FT	METAL 521 521
FT	METAL 683 683
FT	BINDING 672 672
FT	CARBOHYD 136 136
FT	CARBOHYD 231 231
FT	CARBOHYD 293 293
FT	CARBOHYD 617 617
FT	CARBOHYD 665 665
FT	SEQUENCE 762 AA; 84883 MW; BB43D047744AF2 CRC64;
QY	Query Match 14.7%; Score 67; DB 1; Length 762;
Db	Best Local Similarity 38.6%; Pred. No. 24; Matches 22; Conservative 3; Mismatches 18; Indels 14; Gaps 3
QY	22 LWTGL-----GEGGEGGIGPEGQASPTTPDCASRWPSASRWPSAGLTVADRPQL 71
Db	9 LWTLLVMGREGGVSGEGVGKQCHPSLPFRCPSRSP---SDQPW----THPDQSQL 58
DR	RESULT 10
DR	BAI2_HUMAN STANDARD; PRT; 1572 AA.
ID	AC O60241;
DT	DT 16-OCT-2001 (Rel. 40, Created)
DT	DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT	DT 28-FEB-2003 (Rel. 41, Last annotation update)
DR	Brain-specific angiogenesis inhibitor 2 precursor.
GN	BAI2.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID:9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Fetal brain;
RA	MEDLINE=98194217; PubMed=9533023;
RR	Shiratsuchi T., Nishimori H., Ichise H., Nakamura Y., Tokino T.;
RT	"Cloning and characterization of BAI2 and BAI3, novel genes homologous
RT	to brain-specific angiogenesis inhibitor 1 (BAI1).";
RL	Cytogenet. Cell Genet. 79:103-108(1997).
CC	-!- FUNCTION: MIGHT BE INVOLVED IN ANGIOGENESIS INHIBITION.
CC	-!- SUBCELLULAR LOCATION: Integral membrane protein.
CC	-!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN BRAIN. ALSO DETECTED IN
CC	HEART, THYMUS, SKELETAL MUSCLE, AND DIFFERENT CELL LINES.
CC	-!- SIMILARITY: Belongs to family 2 of G-protein coupled receptors.
CC	-!- SIMILARITY: Contains 4 TSP type-1 domains.
CC	-!- SIMILARITY: Contains 1 GSP domain.
CC	This SWISS-PROT entry is copyright. It is produced through a collabor
CC	between the Swiss Institute of Bioinformatics and the EMBL outstat
CC	the European Bioinformatics Institute. There are no restrictions on
CC	use by non-profit institutions as long as its content is in no
CC	modified and this statement is not removed. Usage by and for comm
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/anno">http://www.isb-sib.ch/anno</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	EMBL; AB005298; BAA25362.1; -.
DR	PIR; T00027; T00027.
DR	Genew; HGNC:944; BAI2.
DR	MM; 602683;
DR	InterPro; IPR000932; GPCR_secretin.
DR	InterPro; IPR001879; hormn_receptor.
DR	InterPro; IPR000203; PKD_cys_rich.
DR	InterPro; IPR000884; TSP1.
DR	Pfam; PF00002; 7tm 2; 1.



```
RESULT 12
BGAL_KLEPN
ID BGAL_KLEPN STANDARD; PRT; 1034 AA.
AC P06219;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Beta-galactosidase (EC 3.2.1.23) (lactase).
GN LACZ.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Klebsiella.
OX NCBI_TaxID=573;
RN [1];
RP SEQUENCE FROM N.A.
RX MEDLINE=85289025; PubMed=3897196;
RA Buvinger W.E., Riley M.;
RT "Nucleotide sequence of Klebsiella pneumoniae lac genes.";
RL J. Bacteriol. 163:850-857(1985).
CC -|- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC galactose residues in beta-D-galactosides.
CC -|- SUBUNIT: Homotetramer.
CC -|- SIMILARITY: Belongs to family 2 of glycosyl hydrolases.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M11441; AAA25082.1; -.
DR PIR; A24925; A24925.
DR HSP; P00722; 1BGL.
DR InterPro; IPR008979; Gal_bind_like.
DR InterPro; IPR006101; Glyco_hydro_2.
DR InterPro; IPR006102; Glyco_hydro_21g.
DR InterPro; IPR006104; Glyco_hydro_2SB.
DR InterPro; IPR006103; Glyco_hydro_2TIM.
DR InterPro; IPR004200; Glyco_hydro_42C.
DR InterPro; IPR004199; Glyco_hydro_42N.
DR Pfam; PF02930; Bgal_small_C; 1.
DR Pfam; PF02929; Bgal_small_N; 1.
DR Pfam; PF00703; Glyco_hydro_2; 1.
DR Pfam; PF02836; Glyco_hydro_2_C; 1.
DR Pfam; PF02837; Glyco_hydro_2_N; 1.
DR PRINTS; PR00132; GLYDRLASE2_N; 1.
DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR PROSITE; PS00508; GLYCOSYL_HYDROL_F2_2; 1.
KW Hydrolase; Glycosidase.
FT ACT_SITE 468 468 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 544 544 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 544 544 117517 MW; 8F8B9DC7521EF649 CRC64;
SQ SEQUENCE 1034 AA; 14.4% Score 65.5; DB 1; Length 1034;
Query Match 14.4%; Score 65.5; DB 1; Length 1034;
Best Local Similarity 32.7%; Pred. No. 45;
Matches 17; Conservative 6; Mismatches 20; Indels 9; Gaps 3;
QY 18 HW-----TGLMTGLGEGGIGEGGQASPTPCDCASRWPSASGRWPSAGL 63
DB 983 HHKHKQAEQGWITL-DGLHMGV--GDDSWTPSVLPQWLLSQTRWQYVSL 1031

RESULT 13
CA14_MOUSE
ID CA14_MOUSE STANDARD; PRT; 1669 AA.
AC P02463;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Collagen alpha 1(IV) chain precursor.
GN COL4A1.
OS Mus musculus (Mouse).
OC Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1];
RP SEQUENCE FROM N.A.
RX MEDLINE=89197932; PubMed=2703490;
RA Muthukumar G., Blumberg B., Kurkinen M.;
RT "The complete primary structure for the alpha 1-chain of mouse
RT collagen IV. Differential evolution of collagen IV domains.";
RL J. Biol. Chem. 264:6310-6317(1989).
RN [2];
RP SEQUENCE OF 1-1154 FROM N.A.
RX MEDLINE=88112221; PubMed=3338568;
RA Wood L., Theriault N., Vogeli G.;
RT "cDNA clones completing the nucleotide and derived amino acid
RT sequence of the alpha 1 chain of basement membrane (type IV) collagen
RT from mouse.";
RL FEBS Lett. 227:5-8(1988).
RN [3];
RP SEQUENCE OF 1149-1424 FROM N.A.
RX MEDLINE=86301886; PubMed=3755692;
RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.;
RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a
RT synthetic oligodeoxynucleotide.";
RL Gene 43:301-304(1986).
RN [4];
RP SEQUENCE OF 1276-1669 FROM N.A.
RX MEDLINE=85127033; PubMed=2578961;
RA Oberhaeumer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,
RA Vogeli G., Voss T., Siebold B., Glanville R.W., Kuhn K.;
RT "Amino acid sequence of the non-collagenous globular domain (NC1) of
RT the alpha 1(IV) chain of basement membrane collagen as derived from
RT complementary DNA.";
RL Eur. J. Biochem. 147:217-224(1985).
RN [5];
RP SEQUENCE OF 1441-1669 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RA Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
RT alpha 1(IV) and alpha 2(IV) collagen.";
RL J. Biol. Chem. 262:8496-8499(1987).
RN [6];
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=86196099; PubMed=3009468;
RA Sakurai Y., Sullivan M., Yamada Y.;
RT "Alpha 1 type IV collagen gene evolved differently from fibrillar
RT collagen genes.";
RL J. Biol. Chem. 261:6654-6657(1986).
RN [7];
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89066738; PubMed=3198626;
RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
RT "Head-to-head arrangement of murine type IV collagen genes.";
RL J. Biol. Chem. 263:19274-19277(1988).
RN [8];
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89071759; PubMed=3200851;
RA Burbelo P.D., Martin G.R., Yamada Y.;
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
RT bidirectional promoter and a shared enhancer.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
RN [9];
RP SEQUENCE OF 1-129 FROM N.A.
RX MEDLINE=88243724; PubMed=3379041;
RA Killen P.D., Burbelo P., Sakurai Y., Yamada Y.;
RT "Structure of the amino-terminal portion of the murine alpha 1(IV)
RT collagen chain and the corresponding region of the gene.";
RL J. Biol. Chem. 263:8706-8709(1988).
CC -|- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
```



the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; AB070238; BAB85125.1; -
DR EMBL; AK095734; BAC04618.1; -
DR EMBL; U80738; AAB91437.1; -
DR Genew; HGNC:11955; ZNF384.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zF-C2H2; 8.
DR ProDom; PD000003; Znf_C2H2; 4.
DR SMART; SM00355; Znf_C2H2; 8.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 8.
KW Transcription regulation; Zinc-finger; Metal-binding; Nuclear protein;
KW DNA-binding; Repeat; Alternative splicing.
FT ZN_FING 228 250 C2H2-TYPE 1.
FT ZN_FING 256 278 C2H2-TYPE 2.
FT ZN_FING 284 306 C2H2-TYPE 3.
FT ZN_FING 317 339 C2H2-TYPE 4.
FT ZN_FING 345 367 C2H2-TYPE 5.
FT ZN_FING 373 397 C2H2-TYPE 6.
FT ZN_FING 403 425 C2H2-TYPE 7.
FT ZN_FING 433 455 C2H2-TYPE 8.
FT DOMAIN 461 521 GLN-RICH.
FT DOMAIN 466 499 ALA-RICH.
FT VARSPIC 300 360 Missing (in isoform 2).
FT FTID=VSP_006920.
SQ SEQUENCE 576 AA; 63091 MW; 2A152786C3C46D90 CRC64;
Query Match 14.0%; Score 64; DB 1; Length 576;
Best Local Similarity 29.1%; Pred. No. 36;
Matches 25; Conservative 6; Mismatches 21; Indels 34; Gaps 6;
QY 11 GCGFADAHWTGLMT-----GLG-----EQEGGIGPEGQAS-----PTP----- 44
DB 41 GCGLAPPHYTLTVPASVSLPFGISMDTSKSQLTPHSQASVTQNTVVVPSTGLMT 100
QY 45 ---DCASWRP--SASRWPWSAGLTV 65
DB 101 AGVSCSQWRREGSQSRGP---GLVI 123
RESULT 15
TRYA_HUMAN
ID TRYA_HUMAN STANDARD; PRT; 275 AA.
AC P15157; QSH2Y5; Q9UQL1;
DT 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alpha-tryptase precursor (BC 3.4.21.59) (Tryptase 1).
GN TP51.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Lung;
RX MEDLINE=90009311; PubMed=2677049;
RA Miller J.S., Westin E.H., Schwartz L.B.;
RT "Cloning and characterization of complementary DNA for human
RT tryptase.";
RL J. Clin. Invest. 84:1188-1195 (1989).
RN [2]
RP REVISIONS TO 89-93 AND 108.
RA Schwartz L.B.;
RL Submitted (MAR-1990) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=99121069; PubMed=9920877;

```

```

RA Pallaoro M., Pejzo M.S., Shayesteh L., Blount J.L., Caughey G.H.;
RT "Characterization of genes encoding known and novel human mast cell
RT tryptases on chromosome 16p13.3.";
RL J. Biol. Chem. 274:3355-3362 (1999).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Lung;
RA Wang H.W., McNeill H.P., Thomas P.S., Murphy B.N., Webster M.J.,
RA Hettiaratchi A., King G., Heywood G.J., Huang C., Stevens R.L.,
RA Hunt J.E.;
RT "Molecular cloning and characterization of novel human tryptase cDNAs
RT and splicing variants.";
RN Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 31-50, AND PITUITARY, SEQUENCE OF 31-38.
RC TISSUE=Lung;
RX MEDLINE=87109258; PubMed=3543004;
RA Cromlish J.A., Seidah N.G., Marcinkiewicz M., Hamelin J., Johnson D.A.,
RA Chretien M.;
RT "Human pituitary tryptase: molecular forms, NH2-terminal sequence,
RT immunocytochemical localization, and specificity with prohormone and
RT fluorogenic substrates.";
RL J. Biol. Chem. 262:1363-1373 (1987).
CC -!- FUNCTION: Tryptase is the major neutral protease present in mast
CC cells and is secreted upon the coupled activation-degranulation
CC response of this cell type.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-, Lys-|-, but
CC with more restricted specificity than trypsin.
CC -!- SUBUNIT: Homotetramer.
CC -!- SUBCELLULAR LOCATION: Released from the secretory granules upon
CC mast cell activation.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=P15157-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P15157-2; Sequence=VSP_005374;
CC -!- SIMILARITY: Belongs to peptidase family S1. Tryptase subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

EMBL; M30038; AA86934.1; -
EMBL; AF098328; AAD17846.1; -
EMBL; AF206665; AAG35695.1; -
EMBL; AF206666; AAG35696.1; -
HSSP; P20231; 1AOL.
MEROPS; S01.015; -
MEROPS; S01.143; -
Genew; HGNC:12018; TP51.
MIM; 191080; -
GO; GO:0008236; F-serine-type peptidase activity; TAS.
GO; GO:0006952; P-defense response; TAS.
InterPro; IPR009003; Cys_Ser_trypsin.
InterPro; IPR001254; Peptidase_S1.
InterPro; IPR001314; Peptidase_S1A.
Pfam; PF00089; trypsin; 1.
PRINTS; PR00722; CHYMOTRYPSIN.
PROSITE; PS0240; TRYPsin_DOM; 1.
PROSITE; PS00134; TRYPsin_HIS; 1.
PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Serine protease; Signal; Glycoprotein; Zymogen;
KW Polymorphism; Alternative splicing.
FT SIGNAL 1 18 POTENTIAL.
FT PROPEP 19 30 ACTIVATION PEPTIDE.
FT CHAIN 31 275 ALPHA-TRYPTASE.
FT ACT_SITE 74 74 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 121 121 CHARGE RELAY SYSTEM (BY SIMILARITY).

```

```

FT ACT SITE      224      224      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID      59      75      BY SIMILARITY.
FT DISULFID     155     230      BY SIMILARITY.
FT DISULFID     188     211      BY SIMILARITY.
FT DISULFID     220     248      BY SIMILARITY.
FT CARBOHYD     132     132      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD     233     233      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC      79      87      Missing (in isoform 2).
FT VARSPLIC      79      87      /FTid=VSP 005374.
FT VARIANT      15      15      R -> P (IN ALPHA-II).
FT VARIANT      15      15      /FTid=VAR 012102.
FT VARIANT     221     221      K -> Q (IN ALPHA-II; dbSNP:1137382).
FT CONFLICT     215     216      /FTid=VAR 012103.
FT SEQUENCE     275 AA; 30772 MW; B9BAC4BBCB91CE75 CRC64;
                                TR -> SQ (IN REF. 1).

```

```

Query Match      13.9%; Score 63.5; DB 1; Length 275;
Best Local Similarity 38.5%; Pred. No. 20;
Matches 15; Conservative 3; Mismatches 8; Indels 13; Gaps 2;

```

```

OY      30 QEGGIGPEGQASPTFDCASRWPRGASRWPSAGLTVRDR 68
Db      28 QAGIVG--GQAP-----RSKWPQVSLRVDR 53

```

```

Search completed: August 13, 2004, 09:06:32
Job time : 15 secs

```





```

DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE OSUNBA0088A01.5 protein.
GN OSUNBA0088A01.5
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,
RA Liu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
RA Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
RA Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
RA Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
RA Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
RA Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AL662987; CAD41366.1;
SQ SEQUENCE 813 AA; 90716 MW; 7B50A0B76BF8E5D4 CRC64;

Query Match 17.3%; Score 79; DB 10; Length 813;
Best Local Similarity 33.9%; Pred. No. 10;
Matches 20; Conservative 8; Mismatches 21; Indels 10; Gaps 3;

QY 3 GRPRVSGAGCGF-----ADAHWLGLWTGLGEGQBG-GIGEGQASPTPDCASRWPRGAS 55
Db 2 GDASRIDPSCGSGWRWRCDSHHKALWSSLDGGDAEGSGPDGPA----DGAATWGRSS 56

RESULT 3
Q ID Q9L4W3 PRELIMINARY; PRT; 11096 AA.
AC Q9L4W3
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Nysc.
GN Nysc.
OS Streptomyces noursei.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1971;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 11455;
RX MEDLINE=20334850; PubMed=10873841;
RA Brataset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA Valla S., Zotchev S.B.
RT "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT deduction of the biosynthetic pathway."
RL Chem. Biol. 7:395-403(2000).
DR EMBL; AF263912; AAF71776.1;
DR HSP; P25715; 1MLA.
DR GO; GO:0004024; F:alcohol dehydrogenase activity, zinc-dependent; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0004314; F:[acyl-carrier protein] S-malonyltransferase. .; IEA.
DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001227; Ac trans.
DR InterPro; IPR002065; Adh zn_family.
DR InterPro; IPR000794; FAbD_zn_family.
DR InterPro; IPR006162; Ppantne S.
DR InterPro; IPR006163; Pp bind.
DR Pfam; PF00698; Acyl_transf. 6.
DR Pfam; PF00107; ADH_Zinc_N; 1.

Query Match 17.1%; Score 78; DB 2; Length 11096;
Best Local Similarity 37.2%; Pred. No. 1.6e+02;
Matches 32; Conservative 7; Mismatches 33; Indels 14; Gaps 6;

QY 4 RPRVSGAGCFADAHWTGL---WTGLGEGQBGIGEGQASPTPDCASRWPRGASR--ASR 56
Db 836 REEELSVAVTGLARAHVRGVTVRWAGLFD---GTGAREADLFTYPPQRFQREWPTAARAAQ 891
QY 57 WPWSAGLTVDRDPQLG---ELCMGRG 79
Db 892 DVTAAGLGAADHPLLGATVELADGAG 917

RESULT 4
Q ID Q8TDV3 PRELIMINARY; PRT; 235 AA.
AC Q8TDV3
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative G-protein coupled receptor.
GN GPCR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Takeda S., Kadowaki S., Haga T., Takaesu H., Mitaku S.;
RT "Identification of G protein-coupled receptor genes from the human
RT genome sequence."
RL Submitted (APR-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AB083587; BAB89300.1;
DR GO; GO:0004872; P:receptor activity; IEA.
DR GO; GO:0003700; P:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000847; HTH_Lyxr.
DR PROSITE; PS00044; HTH_LYXR_FAMILY; 1.
RW Receptor.
SQ SEQUENCE 235 AA; 24154 MW; 9B1071B0D7D6B30B CRC64;

Query Match 16.9%; Score 77; DB 4; Length 235;
Best Local Similarity 37.2%; Pred. No. 4.8;
Matches 16; Conservative 3; Mismatches 18; Indels 6; Gaps 1;

QY 19 WTGLWTGLGEGQBGIGEGQASPTPDCASRWPRGASRWPSA 61
Db 11 WLGLWVGL-----GLRFTFRVCSPLCGPLWPSASLCVWGS 47

RESULT 5
Q ID Q84IT5 PRELIMINARY; PRT; 1485 AA.
AC Q84IT5
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Salinomycin polyketide synthase (Fragment).
GN SALA.
OS Streptomyces albus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1888;

```

```

RN  SEQUENCE FROM N.A.
RC  STRAIN-ATCC21838;
RA  Izumikawa M., Murata M., Tachibana K., Ebizuka Y., Fujii I.;
RT  "Cloning of Polyketide Synthase Genes Involved in Salinomycin
RT  Biosynthesis from Streptomyces albus.";
RL  Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AB087998; BAC54914.1; -.
DR  GO; GO:0016740; F:transferase activity; IEA.
DR  GO; GO:0005215; F:transporter activity; IEA.
DR  GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR  GO; GO:0008152; P:metabolism; IEA.
DR  GO; GO:0006810; P:transport; IEA.
DR  InterPro; IPR001227; AC trans.
DR  InterPro; IPR000794; Ketoacyl synth.
DR  InterPro; IPR000566; Lipocin cytfabp.
DR  InterPro; IPR006162; Peantne S.
DR  InterPro; IPR006163; Pp bind.
DR  Pfam; PF00698; Acyl trans; 1.
DR  Pfam; PF00109; Ketoacyl-synt; 1.
DR  Pfam; PF02801; ketoacyl-synt C; 1.
DR  Pfam; PF00550; pp-binding; 1.
DR  PROSITE; PS00075; ACP DOMAIN; 1.
DR  PROSITE; PS00606; B_KETOACYL SYNTHASE; 1.
DR  PROSITE; PS00213; LIPOCALIN; 1.
DR  PROSITE; PS00012; PHOSPHOPANTHINE; 1.
FT  NON_TER 1
FT  NON_TER 1485
SQ  SEQUENCE 1485 AA; 153535 MW; 34AD09A885311B1D CRC64;

Query Match 16.7%; Score 76; DB 2; Length 1485;
Best Local Similarity 33.8%; Pred. No. 36;
Matches 22; Conservative 4; Mismatches 25; Indels 14; Gaps 3;

QY 3 GRPRVSGAGCFADAHWTGLWT-----GLGEGQGGTG-PEGQASPTPCASRWR 52
DB 1053 GRPRISRR-----SAWTRGWSRRFRCPPTGTGRHRRMPGATGSPGCGFRPGR 1108

QY 53 SASRW 57
DB 1109 FEGRW 1113

RESULT 6
Q86SH7 PRELIMINARY; PRT; 205 AA.
AC Q86SH7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=Brain;
RA Strausberg R.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC046165; AAH46165.1; -.
DR EMBL; BC048278; AAH48278.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 205 AA; 22007 MW; 672587E17E2F6485 CRC64;

Query Match 16.2%; Score 74; DB 4; Length 205;
Best Local Similarity 32.1%; Pred. No. 8.5;
Matches 26; Conservative 3; Mismatches 22; Indels 30; Gaps 4;

QY 6 RRVSGAGCFADAHWTGLW-TGL-----GEGQGGIGPEGQASPTPCAS 48
DB 19 RRLGAGSARVQRPAAGRWGLPGEVVVSWLEPRGRSGRAGGGERDGLWRP-----G 73
```

```

QY 49 RWPR-----SASRWPSA 61
DB 74 RWPRGARAFGEQAASRSPSA 94

RESULT 7
Q28936 PRELIMINARY; PRT; 448 AA.
ID Q28936;
AC Q28936;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DE Sus scrofa (Pig).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9623;
RN [1]
RP SEQUENCE FROM N.A.
RA Murakawa M.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=93269219; PubMed=8497848;
RA Murakawa M., Okamura T., Kamura T., Shibuya T., Harada M., Nihō Y.;
RT "Diversity of primary structures of the carboxy-terminal regions of
RT mammalian fibrinogen A-alpha-chains: Characterization of the partial
RT nucleotide and deduced amino acid sequences in five mammalian
RT species.";
RL Thromb. Haemost. 69:351-360(1993).
DR EMBL; D43760; BAA07817.1; -.
DR HSP; P02671; IFZG.
FT NON_TER 1
FT NON_TER 448
SQ SEQUENCE 448 AA; 47355 MW; D49A3CAD1F6F0A92 CRC64;

Query Match 16.2%; Score 74; DB 6; Length 448;
Best Local Similarity 30.6%; Pred. No. 18;
Matches 26; Conservative 2; Mismatches 41; Indels 16; Gaps 4;

QY 3 GRPRVSGAG---CGFADAHWTGLWTGLGEGQGGIGPEGQASPTPCASRW-----PRS 53
DB 129 GRPEPGSTGTWDSGHPDPSAGTWKP-GRPEPGSTGTWDSGHPDPSAGTWKPRPEPGS 187

QY 54 ASRW-----PWSAGLTVDRPQLG 72
DB 188 TGTWDSGRPDPSAGTWKPRPESG 212

RESULT 8
Q86TK2 PRELIMINARY; PRT; 311 AA.
ID Q86TK2;
AC Q86TK2;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Similar to tankyrase 1 binding protein 1, 182kDa (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripheral Nervous System;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC046216; AAH46216.1; -.
FT NON_TER 1
SQ SEQUENCE 311 AA; 33178 MW; 4073CFAP286890F0 CRC64;

Query Match 16.1%; Score 73.5; DB 4; Length 311;
Best Local Similarity 27.6%; Pred. No. 14;
```



```

DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Flageiiform silk protein-like protein.
GN OSJNBA0073D04.40.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoae; Oryza.
OX NCBI_TaxID=39347;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA The Rice Chromosome 10 Sequencing Consortium;
RT "In-depth view of structure, activity, and evolution of rice
RL chromosome 10.";
RL Science 300:1566-1569 (2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Wang R.A., McCombie W.R., Messing J., Yuan Q.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017097; AAP53897.1; -.
SQ SEQUENCE 313 AA; 32336 MW; E749933F20DCE692 CRC64;

Query Match
Best Local Similarity 27.5%; Score 71.5; DB 10; Length 313;
Matches 28; Conservative 5; Mismatches 34; Indels 35; Gaps 6;

QY 2 CGRRVRSAGCGFADAHWTG-----LWTGL-----GEGQGGIG----- 35
Db 26 CGRGGRPAAPARPARHWRGGGLWLVGEASRQRARADRGSSVPVGSRRRRRGGG 85
QY 36 -----PEQASPTPCASRWPSASRPWSAGLTVRDRPQIG 72
Db 86 QLSLPEAATAP-PLTRGAWPGAS--PWRGG-SSRPWRIG 123

RESULT 13
Q80T82 PRELIMINARY; PRT; 597 AA.
AC Q80T82
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MKIAA1858 protein (Fragment).
GN MKIAA1858
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22579291; PubMed=12693553;
RA Okazaki N., Kikuno T., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA Nakajima D., Nakase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RL DNA Res. 10:35-48(2003).
DR EMBL; AK122564; BAC5846.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004672; F:protein kinase activity; IEA.
DR GO; GO:0006468; F:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000719; Prot Kinase.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF000096; zf-C2H2; 3.
DR SMART; SM00355; Znf_C2H2; 3.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00028; ZINC FINGER_C2H2_1; 3.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 2.

```

```

FT NON TER 1
FT NON TER 597
SQ SEQUENCE 597 AA; 64414 MW; 0D928EBD912E3951 CRC64;

Query Match
Best Local Similarity 15.7%; Score 71.5; DB 11; Length 597;
Matches 21; Conservative 7; Mismatches 17; Indels 7; Gaps 2;

QY 26 LGEGQGGIGPEQASPTPCASRWPR---SASRWP---WSAGLTVRDRPQ 70
Db 468 LGLGRHGSVNTGKAPLADPKSSRAPRKQATSRVPPVKSRPSGQSSRAPQ 519

RESULT 14
Q82K61 PRELIMINARY; PRT; 267 AA.
AC Q82K61
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN SAV2543.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12592562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL; AP005031; BAC70254.1; -.
DR InterPro; IPR008996; CytoK_IL1_like.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 267 AA; 27932 MW; 7B10A1B8389C67F8 CRC64;

Query Match
Best Local Similarity 15.6%; Score 71; DB 16; Length 267;
Matches 24; Conservative 7; Mismatches 27; Indels 14; Gaps 4;

QY 7 RVSAGCGFADAHWTGLWTGLGEGQE-----GGIGPEQASPTP-----DCASRWPR3- 53
Db 114 RVAVETGRAPA-WMLVHAMGGQGMGVYDVRVPPQPGASAGNVLVVRDGMPPRGR 172
QY 54 ASRWPKSAGLTV 65
Db 173 VVSGPWQAVITL 184

RESULT 15
Q9U3W6 PRELIMINARY; PRT; 359 AA.
AC Q9U3W6
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE MAB-21.
GN MAB-2 OR CG4746.

```

OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Serano T.L., Pendleton J.D., Rubin G.M.;  
 RT "A reverse genetic screen for genes involved in Drosophila  
 RT development";  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF214524; AAF24503.1; -.  
 DR FlyBase; FBgn0029003; mab-2.  
 DR InterPro; IPR004962; Mab-21.  
 DR Pfam; PF03281; Mab-21; 1.  
 SQ SEQUENCE 359 AA; 41220 MW; DCOB31341ACC4B59 CRC64;

Query Match 15.6%; Score 71; DB 5; Length 359;  
 Best Local Similarity 60.0%; Pred. No. 29;  
 Matches 12; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 39 QASPTPCASRWPRASRWP 58  
 Db 169 QITPAFKCAGLWPRASRWP 189

Search completed: August 13, 2004, 09:07:24  
 Job time : 43 secs